ANGLIA RUSKIN UNIVERSITY

THE COGNITIVE BASIS OF GOAL-DIRECTED BEHAVIOUR

ELLEN LAURA CARROLL

A thesis in partial fulfilment of the requirements of Anglia Ruskin University for the degree of Doctor of Philosophy

Submitted: January 2014

ACKNOWLEDGEMENTS

Life has offered me many guiding influences who have, each in their own ways, contributed to the creation of this thesis. Now reaching towards the end of this PhD, I would like to express my gratitude to those people who have been of particular help and support to me along the way.

To my supervisor, Peter Bright, I sincerely thank you for your valuable advice, guidance, support, and reassurance; it has been a pleasure to learn from someone with such insight and understanding. To the academic and technical staff of the Department of Psychology, Anglia Ruskin University, and not least to my second supervisor, Peter Hills: your help will always be appreciated and I will fondly remember my years in this wholly welcoming and supportive department. I would like to give a special thanks to Anja Lindberg, Anna Jones, and Daragh McDermott who have been tremendous supporters and friends, and to the students who participated in these studies.

Thank you to my irreplaceable friends outside of academia, particularly: to Flice, for reading drafts, for always showing an interest, and for a peaceful space to work; to Verity, my housemate, for looking out for me whilst I was trying to juggle writing-up and full-time work; to Debra, for always being there and for your ability to see the big picture; and to Jade, my beautiful sister, for your encouragement and for being a fantastic study companion.

To my Grandfather, Ronald, thank you for sharing your wisdom concerning, amongst many other things, the importance of courage and hard work. To my devoted parents, Susan and Michael, I wholeheartedly thank you for the unwavering love and financial and emotional support that you have given me; without the two of you, this endeavour would have been impossible. And finally, to Ed, I am forever grateful for your ability to make me smile, for your indescribable way of calming me, for your love and companionship, and for your talent in the kitchen!

ANGLIA RUSKIN UNIVERSITY
ABSTRACT

FACULTY OF SCIENCE & TECHNOLOGY

DOCTOR OF PHILOSPHY

THE COGNITIVE BASIS OF GOAL-DIRECTED BEHAVIOUR

ELLEN LAURA CARROLL

January 2014

An unresolved issue exists in the study of mental ability: which aspect of cognition is responsible for the emergence of psychometric "*g*" (Spearman, 1904, 1927), a general factor that predicts performance in all kinds of cognitive tasks and many important life outcomes? On the basis of themes present in the literature on fluid intelligence (Chapter 1), this thesis explores the relative contributions of processing, storage, and task modelling demands to the recruitment of *g* in task performance. Six experiments are presented which employed two computer-based tasks. The tasks were designed such that the level of demand for processing and storage was separated and manipulated in order to establish their relationship to scores on a standard test of *g*. Task manipulations were implemented in the context of varying the number of distinct verbal chunks in which task instructions were presented, whilst controlling for the presented amount of operative task-relevant information. The findings showed that the recruitment of *g* in task performance was strengthened by the presence (versus the absence) of a requirement to inhibit a prepotent response tendency (Experiments 1, 2, and 3), and by the presence of requirements to inhibit a prepared response and to maintain and update information in working memory (Experiments 4 and 5). However, these effects were observed only when task instructions were presented as four (and not as two) distinct rules. Additional findings showed that reconceptualisation of task requirements—that is, flexibly imposing order on a complex set of instructions thus reducing the number of distinct verbal chunks—was dependent on performance on the test of Spearman's *g* (Experiments 3 through 6). These findings are deemed consistent with a task conceptualisation theory of *g*, with real-time execution demand, particularly inhibition, posing as a risk factor for the recruitment of *g* only when task requirements are maintained in mind as a relatively large number of chunks or rules.

Key words: Spearman's *g*; fluid intelligence; working memory; task modelling; chunking; response inhibition

TABLE OF CONTENTS

I. LIST OF FIGURES

II. LIST OF TABLES

III. LIST OF APPENDICES

COPYRIGHT DECLARATION

THE COGNITIVE BASIS OF GOAL-DIRECTED BEHAVIOUR

ELLEN LAURA CARROLL

January 2014

CHAPTER 1

Theoretical Perspectives on the Cognitive Basis of Spearman's *g*

Overview

This introductory chapter presents the conceptual framework and rationale for the experiments presented herein in which the cognitive basis of fluid intelligence was explored. To begin, Spearman's general (*g*) factor of mental ability (Spearman, 1904, 1927) is introduced alongside behavioural and biological evidence supporting its real-life validity. The main body focuses on the aspects of cognition that are postulated as being involved in *g*, and shows that the cognitive nature of *g* is still very much disputed. First, conclusions drawn from analysing the contents of standard fluid intelligence tests, in order to determine the underlying processes involved, are discussed. Second, information processing speed is deemed too basic to reflect something as complex as intelligence, particularly when viewed against the inconsistent relationship between speed and *g*. Third, proposed mediating factors in the strong relationship between *g* and working memory capacity are reviewed, primarily focusing on executive attention, inhibition, and memory maintenance and retrieval. Last, task modelling, which has been shown to be more important to the recruitment of *g* than the real-time performance demands of the task, is highlighted as a promising direction in the study of *g*. The chapter closes with the research framework which involves a systematic exploration of the aspect(s) of working memory that may be fundamental to the recruitment of *g* in task performance.

Spearman's general factor of mental ability

Historical views on mental ability

People have reflected on mental ability for centuries. Plato (427–347 BC), for example, recognised intellect as an important attribute in his human nature metaphor in which a charioteer, intellect, drives a chariot pulled by two horses, emotion and will. It

was not until the latter half of the 19th century, however, that empirical investigation of *individual differences* in human ability was made. Sir Francis Galton, like many other scientists, distinguished between a general ability that impacts on all mental activities and a number of specific abilities that influence specialised activities such as maths, linguistics, memory, music, and art. Noticing that limits in intelligence differed among university students but were more stable among relatives, Galton (1869) argued for the heritable nature of individual differences in general ability, which he believed reflected the ability for sensory discrimination.

Other early and influential advances in the study intelligence were made by Alfred Binet who established the idea of the intelligence quotient (IQ). In 1904 Binet was commissioned to identify French school students who would benefit from additional educational support. Binet believed that, rather than sensory discrimination, individual differences in intelligence would be best captured by complex cognitive measures such as attention, memory, comprehension, imagination, and suggestibility. He developed a scale of 30 tests which assessed the practical ability to adapt to the environment (Binet & Simon, 1905, 1908), from co-ordinating gaze or reach to stimuli and discriminating lines or weights, through repeating digits or sentences and defining sentences, to complex tests of memory, language, suggestibility, and abstraction. Using IQ, the ratio of mental age to chronological age, this scale classified the intellectual capacity of a person in relation to other people of their age. The idea that this scale could be used to identify above-average, as well as below-average, ability was introduced by Terman (1916) who revised and standardised the Simon-Binet scale for use in American populations, forming the Stanford-Binet scale.

Spearman's *g*

Empirical ways of assessing the *structure* of mental ability have since been designed. Factor analysis, for example, is a method for investigating whether a number of related variables of interest, such as scores on different mental ability tests, can be

accounted for by a smaller number of typically independent, unobservable basic variables or factors. Highly correlated variables represent a single factor whereas weakly correlated variables represent independent factors. Emerging factors, however, are purely statistical constructs; the psychological dimension or function that each factor is deemed to reflect is determined by the researcher on the basis of whether they view the proportion of variance accounted for by that factor as large enough to reflect some true function.

The founder of factor analysis, Charles Spearman, advocated "a 'Correlational Psychology,' for the purpose of positively determining all psychical tendencies, and in particular those which connect together the so-called 'mental tests' with psychical activities of greater generality and interest" (Spearman, 1904, p. 10). In the first experimental series of Spearman's (1904) seminal paper, he correlated three measures of sensory discrimination (visual [light], tactile [weight], and auditory [pitch]) with three measures of intelligence (in-school cleverness and two measures of out-of-school "common sense") in a group of 24 children (aged 10–13 years) attending a village school. In the final experimental series, Spearman correlated auditory discrimination with five measures of school performance—musical talent (as rated by the child's music teacher) and the child's rank in school, ranging between 1 and 33, for four school subjects (the Classics, French, English, and mathematics)—in a group of 33 children (aged 9–13 years) attending a High Class Preparatory School. Across each experiment, Spearman discovered that all of the grades were positively correlated, some more strongly than others, and that these correlations could be accounted for by one general factor. He concluded that "all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (Spearman, 1904, p. 179).

Later, Spearman (1927) administered a battery of over 100 cognitive tasks to a large number of people; in line with his 1904 conclusions, and with Galton's (1869)

predictions, Spearman found that a single general ability or "*g*" factor was common to performance on all mental tests, and that a number of narrower specific ability or "*s*" factors were employed for particular tasks. The *g* factor explained the universal positive correlation observed between the test scores; people that scored well on one type of test tended to score well on others. Specific verbal, numerical, spatial, and mechanical factors explained the stronger correlations between particular tests that were not explained by *g*. Subsequently, group factors (e.g., memory, verbal ability, mathematical reasoning, and spatial visualisation) were added to this two-level hierarchy of mental ability which were particular to correlations that were inadequately captured by *g* or *s* (Spearman, 1938).

Modern methods of factor analysis also support a three-level hierarchy of mental ability. The Cattell–Horn–Carroll (CHC) model, for example, places *g* at the apex (stratum III), a number of broad factors below *g* (stratum II), and a number of narrow ability factors below the broad factors (stratum I). This model is arguably the dominant factor model of intelligence today (McGrew, 2005, 2009), mostly due to the extensiveness of Carroll's (1993, 1997) investigations. His three-stratum theory of cognitive ability resulted from a comprehensive exploratory factor analysis on datasets from 460 studies spanning over 50 years (Carroll, 1993). Carroll's study of the relationships between many aspects of cognition—auditory and visual perception, attention, language, memory, reasoning, creativity, and the speed and accuracy of information processing—established the robustness of "positive manifold" between cognitive test scores, leading Jensen (1997) to describe it as "an inexorable fact of nature" (p. 223).

Stratum II of the CHC model incorporates *fluid intelligence* and *crystallised intelligence* (Cattell, 1941, 1963), and a number of other broad abilities such as short-term memory acquisition and retrieval, long-term memory storage and retrieval, and cognitive processing speed (Horn, 1965; Horn & Blankson, 2005; Horn & Noll, 1997). Crystallised intelligence reflects previously learned skills and knowledge and is best

indicated by tests assessing, for example, general knowledge, vocabulary, and numerical ability. Fluid intelligence, in contrast, reflects novel problem solving and abstract thinking ability, and is best measured by tests that do not rely on learned knowledge but instead involve the perception of complex patterns or relations among geometric figures or among elements within a matrix.

Fluid intelligence is considered to be a better predictor of *g* and has been more reliably associated with the functions of the frontal lobes than has crystallised intelligence (e.g., Duncan, Burgess, & Emslie, 1995). Like other frontal functions, fluid intelligence is more vulnerable to cognitive decline; "relative to the Gc abilities, the Gf abilities decline first, decline over the longest period of adult development, and decline most" (Horn, Donaldson, & Engstrom, 1981, p. 39). Fluid intelligence rises until about age 25 years after which it gradually declines, whereas crystallised intelligence continues to grow across the lifespan (Horn & Cattell, 1966). Both abilities decrease more rapidly in the later years, with fluid intelligence demonstrating an accelerated decline after about age 55 years, and with crystallised intelligence decreasing, to a smaller extent, after about age 65 years (Kaufman & Horn, 1996).

Other theories based on factor analysis emphasise the importance of multiple independent intelligences such as Thurstone's (1938) seven primary mental abilities or Gardner's (1983, 1993, 1999) seven to nine. Guilford's (1967) structure-of-intellect model outlined more than 100 dimensions to intelligence; he contested the existence of a general ability factor because, using his own factor analysis methods, he observed correlations between cognitive tests that were not significantly bigger than zero. It was later suggested, however, that Guilford's results were due to artefacts such as measurement and sampling error and inappropriate selection of tests (Alliger, 1988). Sternberg (1977) argued that factors become psychologically meaningful only when the data is rotated; theories on the structure of mental ability differ, in part, due to the rotation method adopted by the researcher. The subjectivity associated with factor analysis is thus not only evident in the way in which function is assigned to each factor,

but also in the researcher's freedom in the rotation method employed. Nonetheless, the majority of theories admit to the existence of both a "higher order" general factor and "lower order" specific factors (Sternberg, 1985), but differ in terms of which type of factor they lay emphasis on. Indeed, Sternberg and Gardner (1982) concluded: "We interpret the preponderance of evidence as overwhelmingly supporting the existence of some kind of general factor in human intelligence. Indeed, we were unable to find any convincing evidence at all that militates against this view" (p. 321).

Evidence for the existence of *g*

Despite the apparent dispute surrounding the relative importance of this general factor of mental ability, much evidence exists for the real-world validity of *g*. *g* not only exists universally in all ages, sexes, races, and cultures (Jensen, 1998), but it also predicts many important life outcomes, such as social mobility and academic attainment (Colom, Escorial, Shih, & Privado, 2007), occupational attainment (Ree & Earles, 1992), law-abidingness (Gordon, 1997), and even health and survival (Deary, Whiteman, Starr, Whalley, & Fox, 2004). Furthermore, *g* predicts job performance in all types of job and on every dimension of performance (Hunter & Hunter, 1984), and does so better than any other known factor(s) including specific abilities (Hunter, 1983a; Schmidt, Ones, & Hunter, 1992), job knowledge (Hunter, 1986), job experience (Hunter & Schmidt, 1996), and vocational interest and personality (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990). Specific abilities (Hunter, 1983a; Schmidt et al., 1992) as well as the conscientiousness–integrity aspect of personality (Schmidt & Hunter, 1998), however, do add to performance in particular jobs. The direct effect of *g* on job performance is greater when jobs are novel (Hunter, 1986) and complex (Hunter, 1983b).

Leading on from Galton's (1869) notion of "hereditary genius," modern behavioural genetic studies show that *g* is highly heritable (Deary, Johnson, & Houlihan, 2009) and has clear biological correlates. Genetic studies provide strong support for a general

factor over a more modular cognitive system; most of the variance in specific abilities is accounted for by genetic influence on *g* (e.g., Pedersen, Plomin, & McClearn, 1994). Petrill (2012) argues that genetic studies demonstrate the existence of a *g* factor because genetic influences upon cognitive abilities increase (e.g., Plomin, Fulker, Corley, & DeFries, 1997) and become more stable (e.g., Plomin, Pedersen, Lichtenstein, & McClearn, 1994) across the lifespan. The biological correlates of *g* include head size (r = .20) and grey matter volume (r = .30 to .40; Rushton & Ankney, 2009), intracranial volume (r = .40; MacLullich et al., 2002), the regularity of ERP waveforms and P300 amplitudes and latencies (Jaušovec & Jaušovec, 2000), brain glucose metabolic rate (Haier, Siegel, Tang, Abel, & Buchsbaum, 1992), and nerve conduction velocity (r = .26; Reed & Jensen, 1992). (Though more recent findings by Reed, Vernon, & Johnson, 2004, suggest that nerve conduction velocity and *g* are unrelated.)

This abundance of real-world evidence for *g*, however, does not detract from the fact that *g* is a purely *statistical* construct. Although some general aspect of cognition may be responsible for the emergence of *g*, it is equally possible that *g* simply represents a mathematical unity rather than a psychological dimension (Brown & Thompson, 1921). Nonetheless, scientific investigation of the cognitive basis of *g* has received much experimental attention. Spearman (1927) viewed *g* as a kind of mental energy and suggested that it would load most highly onto tests that involve the *eduction of relations and correlates* and *abstractness*. Today, it is still unclear whether *g* reflects a single underlying function, like basic *information processing speed* (e.g., Grudnik & Kranzler, 2001) or *working memory* (e.g., Unsworth, Redick, Lakey, & Young, 2010), or instead reflects the average level of a number of processes that function together to perform complex tasks (e.g., Detterman, 1987, 1992, 1996; Anderson, 1992, 1999; Kranzler & Jensen, 1991). The remainder of this chapter explores the aspects of cognition postulated as being involved in *g*, starting with conclusions drawn from studies adopting methods for analysing the content of intelligence tests.

Content analysis of intelligence tests

Raven's Progressive Matrices

Some researchers have analysed the content of tests with high *g* loadings in order to understand the processes involved, and therefore, to understand *g*. The majority of these studies have examined Raven's Progressive Matrices (Raven, 1962; Raven, Court, & Raven, 1988) which was designed to assess the eduction of relations and correlates noted by Spearman (1927). Each Raven's item comprises a 2 × 2, 3 × 3, or 4 × 4 matrix that contains figures that follow patterns across the columns and down the rows of the matrix, with the bottom right section of the matrix left empty. To complete the matrix, these patterns (or rules) must be determined, and the appropriate response must be selected from an answer bank that comprises five alternative responses.

Evidence for goal maintenance

Carpenter, Just, and Shell (1990) hoped to advance understanding of the nature of individual differences in "analytic intelligence." To this aim, they analysed error frequency patterns, eye-fixation patterns, and participants' verbal accounts of any rules encountered whilst solving each Raven's problem. Analysis of the verbal protocols indicated that five rules (which explained patterns in the matrices) were used to solve the majority (69%) of Raven's problems. The five rules were: (a) *constant in a row* in which the value of an attribute (size, position, or number) remains constant within rows but changes between rows; (b) *quantitative pairwise progression* in which the value of an attribute increases or decreases between columns; (c) *figure addition or subtraction* in which the figure in the first column is added to or subtracted from the figure in the second column to produce the figure in the third column; (d) *distribution of three values* in which three values of an attribute occur across three figures in a row; and (e) *distribution of two values* in which two values of an attribute occur across two figures in a row (the third row is valueless).

Error frequency patterns were dependent on the number and type of rule(s) involved in solving a problem. Error rates increased as the number of rules (and therefore, perhaps, the number of attributes held in working memory) associated with solving a problem increased. Any of the five rules were solved by participants that had a high overall Raven's score, but problems involving the distribution of two values rule tended to be failed by participants that had a median (or lower) overall score. Verbal protocols and eye-fixation patterns revealed that problems were solved incrementally; participants described rules consecutively and repeatedly directed their gaze back and forth between two figures (demonstrating a paired comparison). From this initial experiment, Carpenter et al. (1990) reasoned that goal management—"the ability to generate subgoals in working memory, record the attainment of subgoals, and set new subgoals as others are attained" (p. 413)—predicted individual differences in Raven's performance.

To test this theory, Carpenter et al. (1990) then correlated Raven's scores with performance on the Tower of Hanoi puzzle using a goal recursion strategy (Kotovsky, Hayes, & Simon, 1985; Simon, 1975). In the Tower of Hanoi, three or more disks of increasing size are arranged, in the shape of a pyramid, on one of three pegs (the start peg). To solve the puzzle, the pyramid must be moved to another peg (the goal peg) without moving more than one disk at a time and without placing a larger disk onto a smaller disk. The goal recursion strategy involves planning goals and sub-goals prior to performance. In a three disk problem, for example, the first goal involves moving the largest disk to the goal peg. This goal encompasses a number of sub-goals in the order of: (a) moving the smallest disk from the start peg to the goal peg; (b) moving the medium-sized disk from the start peg to the spare peg; (c) moving the smallest disk from the goal peg to the spare peg; and (d) moving the largest disk from the start peg to the goal peg. However, the second goal (i.e., moving the medium-sized disk to on top of the largest disk) and the third goal (i.e., moving the smallest disk to on top of the medium-sized disk) do not have any associated sub-goals. Thus, as the number of disks

involved in a problem increases, so does the number of goals (as well as the number of sub-goals associated with each goal) involved in solving the problem.

Goal recursive Tower of Hanoi performance was indeed strongly[1] correlated (r = .77) with Raven's performance. Furthermore, the number of sub-goals associated with each goal influenced the amount of Tower of Hanoi error in participants with low, but not high, overall Raven's scores. On the basis of their behavioural observations, Carpenter et al. (1990) proposed that individual differences in Raven's performance is determined by the ability to induce abstract relations and to manage multiple problem-solving goals in working memory. They represented this theory as two computer simulation models: BETTERRAVEN knew every rule and mirrored the performance of high Raven's scorers whereas FAIRRAVEN knew all but one of the rules (i.e., all rules except for the distribution of two values rule) and mirrored the performance of median scorers.

The idea that Raven's performance (and therefore *g*) reflects the ability to maintain goal-relevant information (and to inhibit goal-irrelevant information) is not dissimilar to conclusions made by other researchers. Conway, Cowan, Bunting, Therriault, and Minkoff (2002), for example, argue that "the ability to maintain goal-relevant information (i.e., rules) in the face of concurrent processing (i.e., searching for new rules) and distraction (i.e., filtering of irrelevant features) is essential for successful performance on RAVENS" (p. 179). However, the factors associated with goal maintenance, such as the number of rules involved in solving a problem (Unsworth & Engle, 2005) and item difficulty (which is related to the number of rules; Salthouse, 1993), do not affect the relationship between *g* and working memory capacity.

[1] Throughout this thesis, the strengths of correlation coefficients are described using Cohen's (1988) system: a correlation of $r = \pm.50$ is considered strong; a correlation of $r = \pm.30$ is considered moderate; and a correlation of $r = \pm.10$ is considered weak.

Evidence for executive attention

Instead of goal maintenance, other studies emphasise the importance of executive function, or the ability to control attention, in Raven's performance, especially in conditions of distraction. The *focus of attention* is viewed as the aspect of working memory that maintains several items in awareness and protects the items from decay and interference (Cowan, 1988, 1995, 2005). Wiley, Jarosz, Cushen, and Colflesh (2011), for example, rejected the hypothesis that the relationship between working memory capacity and Raven's Advanced Progressive Matrices is mediated by some storage aspect of working memory and instead compared the plausibility of the ability to learn rules quickly (e.g., Guthke & Stein, 1996; Verguts & De Boeck, 2002) and the ability to resist distraction (e.g., Darowski, Helder, Zacks, Hasher, & Hambrick, 2008). They found that working memory capacity was more strongly correlated with performance on items that required new combinations of rules relative to items involving previously encountered rules, which, they argued, supported the role of executive attention in Raven's performance (the learning efficiency account would predict the reverse pattern). Other researchers suggest that greater executive attention ability or focus of attention, may (a) result in a greater number of items (current and previous) being compared, which would increase the likelihood of the recognition of new patterns among figures (indicating new rules; Oberauer, Süß, Wilhelm, & Sander, 2007), and (b) reduce the likelihood of previously encountered rules interfering with the construction of new rules (Bunting, 2006; May, Hasher, & Kane, 1999).

The role of working memory capacity in performance of Raven's has also been shown to vary as a function of both the *level* of distraction associated with alternative responses and the *type* of Raven's problem. Jarosz and Wiley's (2010) study was based on the assumption that the most commonly selected incorrect response would be more salient than other distractors and would therefore require additional attentional control to ignore. They manipulated the saliency of alternative responses in the answer bank of the Raven's and included the most commonly selected incorrect response in one version

of the task but not another. Despite performance scores being only slightly lower, the correlation between working memory capacity and Raven's performance was stronger, for the task that comprised the most salient incorrect response. Wiley and Jarosz (2012) built on these findings by showing that working memory capacity is more strongly related to problems in which a novel, rather than known, rule combination is required. However, they were unsure of whether this is due to the ability to resist interference from previously encountered rules, the ability to attend to multiple elements of a problem at once, or the accessibility of rules in memory.

Strategy Use

Intelligence test problems also differ in terms of the type of strategy involved in solving them. Several published studies have made distinctions between items that rely on verbal-analytical reasoning and those that rely on visual perception (e.g., Hunt, 1974; Dillon, Pohlmann, & Lohman, 1981; Hertzog & Carter, 1988). DeShon, Chan, and Weissbein (1995) provided evidence for distinct visuo-spatial and verbal-analytic strategies in (a) participants' verbal reports (both concurrent and retrospective) of solving each Raven's problem, and (b) the emergence of a verbal overshadowing effect in which describing the problem hindered performance on visuo-spatial items only. A number of studies (e.g., van der Ven & Ellis, 2000; Lynn, Allik, & Irwing, 2004; Gimmig, Huguet, Caverni, & Cury, 2006) have shown that verbal-analytic items, which rely on sequential, hypothetical processing (Lynn et al., 2004; McGrew & Flanagan, 1998), are a better predictor of working memory capacity than items that can be solved on the principle of some Gestalt continuation rule such as similarity or good continuation. According to Gestalt theory, this occurs without conscious effort by an innate disposition to perceive patterns in the world (Kellman, 2000).

Similarly, Raven's items have been dissociated in terms of whether they involve either *constructive matching* or *response elimination*. That is, constructing the answer by viewing the matrix and comparing this potential response to the response

alternatives, or, eliminating incorrect answers (and arriving at the correct answer by default) by comparing the features of each alternative response against the features of the matrix (Bethel-Fox, Lohman, & Snow, 1984). Evidence indicates that constructive matching is used to solve easier Raven's items whilst response elimination is adopted as item difficulty increases, and that people who score highly on tests of intelligence tend to persist with constructive matching for longer (Bethel-Fox et al., 1984; Vigneau, Caissie, & Bors, 2006). This is probably because working memory demand, and therefore possibly reliance on *g*, is greater in constructive matching because it requires top-down processing whereas response elimination is perceptually guided.

The most established strategy distinction explaining performance on complex tasks is speed versus accuracy. Evidence for a relationship between *g* and individual differences in speed-accuracy preferences is not compelling. Speed-accuracy preferences are generally, but not always, stable across different tests (Phillips & Rabbitt, 1995), but they can also be dependent on the specific demands of the task (Jones & Duffy, 1982). Furthermore, when performance is time-limited, two individuals with the same total score could favour alternate strategies; a speed preference results in more questions being answered but with some errors, whereas an accuracy preference results in no-to-few errors but fewer answered questions. It is more plausible that any individual differences in speed-accuracy preferences reflect personality rather than ability. Impulsive individuals and extroverts prefer speed whereas reflective individuals and introverts prefer accuracy (Jensen, 1982; Dickman & Meyer, 1988; see Phillip & Rabbitt, 1995, for conflicting evidence). Generally, though, fast performers tend to perform more accurately (Rabbitt, 1990).

Speed of information processing and *g*

Mental speed

In line with Galton's notion that mental ability reflects *synaptic efficiency*, a number of authors have explained individual differences in *g* on the basis of individual

differences in the speed of information processing. In Anderson's (1992, 1999) theory of minimal cognitive architecture, for example, speed of information processing accounts for a large proportion of the variance in general intellectual ability. In everyday conversation, high performers are commonly described as "quick-witted" or "fast at grasping things," and poor performers are often described as "slow." Slow processing could indeed render it more difficult to handle complex pieces of information because less information can be acquired (Vernon, 1983).

Moderate, albeit sometimes variable (Deary & Caryl, 1997), correlations have also been observed between IQ and a number of biological measures assumed to represent speed of information processing, including evoked potential measures (Caryl, 1994), peripheral nerve conduction velocity (Vernon & Mori, 1992), and eye-blink reflex modification (Smyth, Anderson, & Hammond, 1999). Other evidence suggests that speed of information processing is faster in "gifted" children and in individuals with "higher aptitude" (Cohn, Carlson, & Jensen, 1985; Jensen, Cohn, & Cohn, 1989; Kranzler, Whang, & Jensen, 1994; Rabbitt & Goward, 1994; Saccuzzo, Johnson, & Guertin, 1994) and slower in mentally handicapped individuals (e.g., Baumeister & Kellas, 1968). Furthermore, the relationship between age and scores on tests of fluid intelligence mirrors the relationship between age and processing speed (Sheppard & Vernon, 2001).

Reaction time measures

Hick's (1952) Choice Reaction Time task is a classic measure of mental speed that has been used extensively in research focussed on the relationship between mental speed and *g*. The Hick test involves a small electronic console which comprises eight light buttons that are situated around the edge of a home button. The participant places a finger on the home button, and, after a warning signal (followed by a variable interval of one to four seconds), one of the eight lights turns on. The participant must turn the light off as quickly as possible by moving their finger onto the relevant light button.

The task generally involves four blocks of 15 trials in which the number of visible buttons increases from one, through two and four, to eight (the remaining buttons are shielded from view using a cover). The time between the release of the home button and the pressing of the appropriate response button, or *movement time*, stays quite constant regardless of number of choices. However, the time between the light turning on and the home button being released, or *reaction time*, increases as the number of choices increases because the number of comparisons that need to be computed is greater (this is known as Hick's law).

Many studies report that speed of information processing, as measured by the rate of increase in reaction time with a greater number of choices, is correlated with individual differences in *g*, although the correlation is generally only weak to moderate in strength (r = -.20 to -.40; Jensen, 1982, 1998; Sheppard, 2007). This correlation between reaction time and *g* increases (to around r = -.50) when a battery of choice reaction time tasks is used, and when the complexity of the Hick test is increased via the inclusion of a dual task (Jensen, 1998) or competing task (Fogarty & Stankov, 1995). Generally, it is the differences in the slower, and not the faster, reaction times that are correlated with intelligence (this is known as the worst performance rule; e.g., Larson & Alderton, 1990; Coyle, 2003); the fastest reaction times of people that score poorly on intelligence tests are almost as fast as the fastest reaction times of high performers, but the slowest reaction times of high performers are rarely as slow as the majority of reaction times of poor performers (Jensen, 1987). A strong criticism of associating intelligence with reaction time measures is the possibility of strategy-use (e.g., trading accuracy for speed when releasing their finger from the home button; Deary & Stough, 1996).

Inspection time measures

Another measure of processing speed—*inspection time*—is more strongly correlated with *g* than reaction time (Jensen, 1998, 2006). Inspection time is the minimum exposure duration required to meet a criterion performance level (which is usually set at

a level of 85% correct) in discriminating between perceptual stimuli. Inspection time tasks involve, for example, deciding which of two vertical lines, joined together by a horizontal line at the top, is longer, or deciding which ear auditory information is presented to. Discrimination of the lengths of the lines is easy at long exposures, but gets more difficult as exposure time shortens. These tasks are arguably a better measure of mental speed than reaction time tasks because they do not involve a speeded response and therefore do not afford the adoption of speed-accuracy trade-offs (Nettelbeck & Lally, 1976; Deary & Stough, 1996). Correlations of r = -.45 (31 studies, N = 1,120; Kranzler & Jensen, 1989) and r = -.51 (92 studies, N = 4,197; Grudnik & Kranzler, 2001) between inspection time and IQ have been reported in meta-analyses.

Sheppard and Vernon (2008) reviewed data from 172 studies (N = 53,542) and reported correlations between general, fluid, and crystallised intelligence and a number of mental speed measures. Weak correlations with fluid intelligence were observed for reaction time (r = -.20 to -.26) and inspection time (r = -.26 and -.29), but a more moderate correlation was observed for general speed of processing (r = -.35) which was measured by tasks that involved, for example, mental arithmetic or sorting letters/numbers into ascending order. Correlations with fluid intelligence were weaker for (a) the speed of retrieval from short-term memory (r = -.15) which was measured by tasks that involved indicating whether a letter or digit shown in the current set was shown in an earlier set, and (b) the speed of long-term memory retrieval (r = -.16) which was measured by, for example, tasks that involved indicating whether stimuli were related in a physical or a semantic way, or whether words were synonyms or antonyms. Overall, the mean correlation between mental speed and intelligence was r = -.24, supporting the small role of processing speed in individual differences in intelligence. However, stronger correlations have been reported elsewhere (e.g., Vernon & Weese, 1993). When reviewing 13 behavioural genetic studies, Sheppard and Vernon (2008) found evidence for genes that contribute to individual differences in

both processing speed and intelligence; they also found a mean genetic correlation between processing speed and intelligence of $r = .73$.

Can *g* be explained by mental speed?

Despite the evidence in support of a relationship between processing speed and intelligence, the idea that processing speed and intelligence are synonymous has been robustly challenged in the literature. Neither reaction time tasks nor inspection time tasks are assumed to involve complex cognitive operations and it seems counter-intuitive to brand something as complex as intelligence as reflecting such a basic and simple physiological function as processing speed. The strength of correlations between processing speed and *g* is dependent on a number of other factors such as increased practice (Ackerman, 1988), strategy use (Grundick & Kranzler, 2001; both reduce correlations with *g*), and task complexity (increases correlations with *g*; Jensen, 1998). The existence of practice effects on processing speed tasks (Logan, 1988) also suggests that performance might not always reflect optimum neural transmission rate but may instead reflect some aspect of attention. Carlson (1985), for example, suggests that higher-level processes like sustained attention may mediate the relationship between *g* and speed given that the standard deviation of reaction time is more related to *g* than the mean (e.g., Baumeister, 1998; Jensen, 1992; Walhovd & Fjell, 2007).

Certainly, many researchers link the relationship between processing speed and *g* with working memory. Vernon (1983), for example, suggests that the need to process information quickly increases as working memory demand increases (due to information becoming more complex) because the working memory system is limited in capacity and the information stored or maintained in working memory decays, or is replaced, over time. Some authors (Fry & Hale, 1996; Kail & Salthouse, 1994) instead argue that processing speed accounts for strong relationships reported between working memory capacity and fluid intelligence; this working memory–*g* relationship is the focus of the next section of this chapter. In sum, although basic information processing

speed is related to intelligence, this relationship is often variable and dependent on other factors, which questions the fundamental importance of processing speed in *g*.

Working Memory and *g*

Working Memory

Working memory, like intelligence, is viewed as a domain general system that is involved in the performance of complex tasks. A wealth of evidence exists linking working memory with intelligence; latent variable analyses have uncovered strong correlations of $r = .50$ to .90 between performance on tests of working memory capacity and performance on tests of fluid intelligence (e.g., Ackerman, Beier, & Boyle, 2002; Conway et al., 2002; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; Unsworth et al., 2010). Other studies have found strong links between working memory and both reasoning ability (Kyllonen & Christal, 1990) and a *g* factor that combined fluid intelligence, crystallised intelligence, spatial ability, and psychometric speed measures (Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004). Evidence also shows that working memory and fluid intelligence share common genetic variance (Luciano et al., 2001). Furthermore, functional magnetic resonance imaging (fMRI) studies of the human brain suggest that the dorsolateral prefrontal and anterior cingulate cortices are associated with both working memory (Smith & Jonides, 1999) and fluid intelligence (Kane & Engle, 2002), and that the activations associated with the two constructs are correlated at $r = .54$ (Gray, Chabris, & Braver, 2003).

Working memory is conceptualised as a kind of mental workspace associated with the concurrent storage and processing of information: "Working memory is assumed to be a temporary storage system under attentional control that underpins our capacity for complex thought" (Baddeley, 2007, p. 1). The original working memory model (Baddeley & Hitch, 1974) described a multicomponent system that comprises functionally separate but closely interlinked parts which mediate between perception, long-term memory, and action. Recent conceptualisations of working memory

(Baddeley, 2012) still incorporate the originally proposed central executive and storage buffers (the visuo-spatial sketch pad [for visual and spatial information] and the phonological loop [for verbal information]). However, Baddeley (2012) now speculates the existence of separate storage subcomponents for haptic, tactile, and kinaesthetic information (in the sketch pad), and music and environmental sounds (in the phonological loop). The central executive system, which primarily involves the control of action, has similarly been fractionated into a number of *executive functions* such as focussing and dividing attention, inhibiting irrelevant information, and task switching. The episodic buffer, which was added to the model more recently (Baddeley, 2000), relies on conscious awareness to temporarily store and integrate information from long-term memory and the storage or *slave systems*.

Baddeley (1986) incorporated the notion of a supervisory attentional system into his conceptual framework of the central executive. According to Norman and Shallice's (1986) theory of *attentional control*, routine actions are controlled by schemas using *contention scheduling*, which is quick and automatic, and novel actions are controlled by an executive monitoring system—*the supervisory attentional system*—which is slow and controlled. When an activity is routine, contention scheduling activates a schema that specifies a series of actions relevant to the task and inhibits competing but inappropriate schemas. When an activity is novel, or when inhibition of habitual responses is crucial to appropriate action, the supervisory attentional system manipulates schema activation probabilities and, in turn, controls contention scheduling, modifies pre-existing schemas (both general and specific) to fit with current goals, and constructs, tests, and implements new schemas when appropriate.

Despite the existence of strong support for a link between fluid intelligence and working memory, the mediating factor(s) in this relationship is not yet agreed. Uncertainty surrounding the source of individual differences in working memory capacity compounds this debate. Individual differences in working memory capacity are attributed to three main factors—executive attention, inhibition, and memory

maintenance and retrieval—which each explain performance on complex tasks, such as the Stroop (Stroop, 1935), slightly differently. In the Stroop task, participants name the ink colour of colour words which are sometimes incongruent with the ink colour (e.g., the word *red* printed in yellow). From the executive attention view (e.g., Engle & Kane, 2004; Kane, Conway, Hambrick, & Engle, 2007), high working memory capacity is associated with a greater capacity for directing attention in a goal-directed fashion (to the task of naming the ink colour) and resolving the response competition between the inappropriate but salient response (reading the word) and the appropriate response (naming the colour). By stressing the importance of inhibitory processes, the inhibition view (e.g., Hasher & Zacks, 1988; Lustig, May, & Hasher, 2001; May et al., 1999) posits that high working memory capacity reflects a greater capacity for restricting working memory access only to task-relevant information and resolving response conflict (consistent with the executive attention account), but also for inhibiting dominant but inappropriate responses (reading the words). The memory maintenance and retrieval view (e.g., Unsworth & Engle, 2007a, 2007b; Unsworth & Spillers, 2010) suggests that individuals with high working memory capacity are better at maintaining relevant information (naming the colour) in working memory, and at avoiding the use of inappropriate retrieval cues (for retrieving the inappropriate response of reading the word from long-term memory).

Employing two versions of a go/no-go task, Redick, Calvo, Gay, and Engle (2011) compared the executive attention, inhibition, and memory maintenance and retrieval theories of working memory capacity. The *simple* version of the task required a response (go) to a target letter (X) and no response (no-go) to non-target letters (non-X; with a reverse mapping in another block), whereas the *conditional* version required a response to target letters (M and W) conditional on the current target differing from the last. The results showed that working memory capacity (as measured by operation, symmetry, and running letter span tasks) differences were only observed in performance of the conditional, and not the simple, task. In the conditional task only, (a) people with

high working memory capacity performed better than people with low working memory capacity on both target trials (e.g., M following W) and lure trials (e.g., M following M), and (b) task performance was significantly correlated with working memory capacity.

Redick et al. (2011) interpreted these findings on the basis of the memory maintenance and retrieval theory of working memory capacity because, they argue, only the conditional task requires the active monitoring and updating of stimulus–response mappings and the retrieval of appropriate goal-relevant responses. The authors reason that the executive attention and inhibition accounts also predict working memory capacity differences in the simple task due to the infrequent presentation of non-target trials; a low presentation rate results in little opportunity for both enforcing the goal-relevant task of ignoring non-target trials and resolving the conflict between the inappropriate but prepotent go response and the appropriate no-go response. Redick et al. (2011) concluded "the ability to rapidly update information within working memory, maintain this information in active memory to guide future behavior, and retrieve this information from inactive memory as needed, are all critical aspects of individual differences in WMC [working memory capacity]" (p. 323) "especially in interference-rich conditions" (p. 308).

Other research demonstrates a link between the level of response competition present in a task and the efficiency of retrieval from long-term memory. Conway and Engle (1994), for example, showed that the relationship between working memory capacity and the speed and accuracy of retrieval from long-term memory is dependent on level of response competition. Perhaps, then, some interaction between inhibitory processes and maintenance/retrieval processes is fundamental to individual differences in working memory capacity. Other evidence suggests that the ability to inhibit prepotent responses, as measured by antisaccade tasks (Roberts, Hager, & Heron, 1994; Mitchell, Macrae, & Gilchrist, 2002) and motor response inhibition tasks (Hester & Garavan, 2005), declines with increasing working memory load. Indeed, Redick et al.'s

(2011) findings could alternatively be explained on the basis of a shared but limited resource for the separate but interlinked components of working memory (e.g., Barrouillet, Bernardin, & Camos, 2004); if resources are used for maintaining the previous target in working memory (i.e., maintaining M or W in the conditional go/no-go task), then fewer resources are available to resolve the conflict associated with, and inhibit the response to, no-go trials.

The remainder of this section focusses on the mediating factors in the relationship between fluid intelligence and working memory capacity, paying particular attention to the three functions postulated as determining inter-individual differences in working memory: executive attention, inhibition, and memory maintenance and retrieval.

Working memory processing and *g*

Some authors argue for the central role of processing (over storage) in driving the correlation between working memory capacity and *g*. This view was borne from the observation that fluid intelligence is better predicted by *complex span*, relative to *simple span*, measures of working memory capacity (e.g., Conway, Kane, & Engle, 2003; Daneman & Carpenter, 1980; Engle, Tuholski, Laughlin, & Conway, 1999; Turner & Engle, 1989), which involve processing, in addition to storage, of information. In typical simple span tasks, participants simply recall stimuli (e.g., letters, words, digits, shapes, or spatial locations presented at a fixed rate) in the correct order, but in complex span tasks, the information to be recalled also needs to be manipulated. In the Reading Span (Daneman & Carpenter, 1980), for example, participants read a series of sentences (processing) and recall the last word from each (storage), and in the Symmetry Span (Kane et al., 2004) participants judge the symmetry of letters or spatial patterns (processing) and, for example, recall, in serial order, the spatial location of another object (storage). A thorough meta-analysis shows that working memory capacity predicts performance on the Raven's regardless of the modality that the complex span task is presented in (Ackerman et al., 2005).

In support of the dissociation between simple and complex span tasks, and therefore separate storage and processing components of working memory, latent variable research has shown that performance on each task loads onto separate latent factors (e.g., Engle et al., 1999; Conway et al., 2002). Only complex span tasks are capable of predicting diverse cognitive abilities including reasoning, reading, listening and language comprehension, vocabulary learning, writing, note taking, following instructions, bridge playing, and computer programming (in addition to fluid intelligence; Engle, 2001). Perhaps simple span tasks rely on automatic schemas whereas the more novel complex span tasks rely on controlled processing, the central executive or supervisory attentional system, with the latter component being more associated with fluid intelligence.

Executive attention was one of the first aspects of working memory postulated as mediating the relationship between working memory capacity and fluid intelligence. Engle et al. (1999), for example, found that attentional control, or executive functioning, matters more to performance on tests of novel problem solving than storage; working memory capacity predicted performance on the Raven's even after individual differences in short-term memory were accounted for. Cowan et al. (2005), like many other researchers, considered the relationship between working memory capacity and fluid intelligence alongside attention. They found that scope of attention constructs and more typical working memory capacity constructs made independent contributions to performance on the Raven's, and thus argued that their scope of attention task measured a different form of basic processing to simple span tasks. Participants briefly viewed a visual array (comprising several squares) and were required to identify any changes in a second visual array; scope of attention was calculated as the highest number of squares that could be accurately processed simultaneously, rather than the number of separately presented items remembered as measured by span tasks. An important function of attention is the ability to focus attention on relevant items and not on irrelevant items and much literature also exists on the relationship between the ability to resist

distraction, working memory capacity, and *g*. Darowski et al. (2008), for example, demonstrated that the level of distractibility (as measured by the time taken to read only the italicised words in a passage comprising both italicised and non-italicised words) is associated with both poorer complex span performance and poorer Raven's performance.

The evidence for the importance of the attentional and executive processes in the performance of fluid intelligence tests is clear. However, an important question yet to be answered is whether a collection of executive functions act together to support fluid intelligence (in line with the idea that *g* reflects the average functioning of a number of cognitive processes), or whether one working memory processing function is more fundamental to individual differences in *g* than the others. Baddeley admits:

> It is probably true to say that our initial specification of the central executive was so vague as to serve as little more than a ragbag into which could be stuffed all the complex strategy selection, planning, and retrieval checking that clearly goes on when subjects perform even the apparently simple digit span task. (Baddeley, 1996, p. 6)

Indeed, the specific roles of executive functions in co-ordinating and controlling complex cognition and action was an "embarrassing zone of almost total ignorance" (Monsell, 1996, p. 93) until about 15 years ago, particularly when viewed against the fast and vast growth of theories on other aspects of cognition such as memory and perception. The neglected question of "unity and diversity" (Teuber, 1972, p. 615) in executive (e.g., Miyake et al., 2000), or frontal (e.g., Duncan, Johnson, Swales, & Freer, 1997), function has since received much experimental attention.

Baddeley (2012) now splits the central executive into several dissociable functions which each have a unique role in the control of action. Individual differences studies also provide evidence for the diversity of executive functioning. These studies typically employ correlation–regression or exploratory factor analyses to examine the relationships between scores across a large battery of complex tasks (including the

Tower of Hanoi, Wisconsin Card Sorting, complex span tasks, and random-number generation) in adults (e.g., Lehto, 1996), older adults (e.g., Lowe & Rabbitt, 1997), and adults with brain damage (e.g., Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Duncan et al., 1997). A consistent observation is that the correlations between the executive tasks are generally weak and non-significant suggesting that a number of separate factors, rather than a unitary factor, explains performance. This view is supported by lesion studies showing that whilst one person will fail one executive task but succeed on another, another person will show the reverse pattern (e.g., Godefroy, Cabaret, Petit-Chenal, Pruvo, & Rousseaux, 1999).

A problem, however, with fractionating executive functioning into specific processes, is just what each distinct factor represents; Miyake et al. (2000) point out that the "WCS [Wisconsin Card Sorting], for example, has been suggested by different researchers as a measure of 'mental set shifting,' 'inhibition,' 'flexibility,' 'problem solving,' and 'categorization,' just to name a few" (p. 53). In an attempt to empirically fractionate executive processing using latent variable analysis, Miyake et al. (2000) discovered three separable (but correlated) executive functions: *inhibition* of prepotent but inappropriate responses, *updating/manipulating* working memory representations to reflect only task-relevant information, and *shifting* between the performance of multiple tasks, mental sets, or operations. They found that each function differentially predicted performance on a range of cognitive tasks. Duncan et al. (1997) observed weak correlations among executive tests scores in individuals with head injury which, they argue, supports the existence of dissociable executive sub-functions such as *switching* (shifting) and *impulse control* (inhibition). However, they also found evidence for a more general function, *g*, which was related to performance on all tasks, indicated by *goal neglect* (i.e., the inability to attend to a known task requirement), and reflected "a process of forming an effective task plan by activation of multiple goals or action constraints" (Duncan et al., 1997, p. 714). Despite these findings, much research

interest has been directed towards the relationship between fluid intelligence and each postulated executive function, particularly inhibition.

Inhibition and *g*

Some authors argue that "intelligence cannot be understood without reference to inhibitory processes" (Dempster, 1991, p. 157). An early theoriser in the field, Louis L. Thurstone, posited that "intelligence, considered as a mental trait, is the capacity to make impulses focal at their early, unfinished stage of formation. Intelligence is therefore the capacity for abstraction, which is an inhibitory process" (Thurstone, 1924/1973, p. 159). A number of studies have indeed shown that psychometric intelligence is significantly correlated with performance on tasks considered to measure response inhibition such as the Stroop (e.g., Dempster, Corkill, & Jacobi, 1995; Polderman et al., 2009; Salthouse, Atkinson, & Berish, 2003), antisaccade (Michel & Anderson, 2009; Unsworth et al., 2010; Unsworth, Spillers, & Brewer, 2009), flanker (Unsworth et al., 2009), negative priming (Borkowski, 1965), and unwanted thought suppression (Brewin & Beaton, 2002; Rutledge, Hollenberg, & Hancock, 1993), tasks. Imaging and genetic findings are at least consistent with the idea that response inhibition and *g* are related. fMRI studies show that the anterior cingulate cortex is recruited in performance of both tasks involving response conflict (e.g., Kerns et al., 2004) and tasks with high *g*-loadings such as Raven's Advanced Progressive Matrices (e.g., Gray et al., 2003). Genetic studies show how both inhibitory control and IQ—as measured by the Wechsler Intelligence Scale for Children-III (Wechsler, 1991) or the Wechsler Adult Intelligence Scale-III (Wechsler, 1997)—also share some genetic factors (e.g., Polderman et al., 2009).

However, non-significant correlations have also been reported between fluid intelligence and performance on, for example, the Stroop (Jensen, 1965; Jensen & Rohwer, 1966; Friedman et al., 2006), proactive and retroactive interference (Jensen, 1964, unpublished; although, conversely, both proactive interference and retroactive

interference were significantly correlated with college grade-point average), and stop-signal (Salthouse et al., 2003) tasks. Studies employing factor analysis have also reported inconsistent observations suggesting that the relationship between inhibition and intelligence is not at all straightforward or consistent. Salthouse et al. (2003) reported a strong relationship (r = .73) with fluid intelligence for their composite measure of inhibition, whereas Friedman et al. (2006) found a non-significant relationship between inhibition and fluid intelligence constructs[2]. In fact, Salthouse et al. (2003) also found that each of their three distinct executive functioning constructs (inhibition, updating, and shifting or "time sharing") was related to fluid intelligence in aging adults, with stronger correlations observed for inhibition and shifting than for updating. But Friedman et al. (2006) found that only updating (and not inhibition or shifting) was strongly correlated with all intelligence measures (i.e., fluid intelligence, crystallised intelligence, and Wechsler Adult Intelligence Scale IQ), suggesting that both working memory and intelligence tests "involve updating to maintain relevant information in the presence of interference" (p. 173). Other studies have also found little relation between intelligence and inhibition (e.g., Chuderski, Taraday, Nęcka, & Smoleń, 2012) or shifting (e.g., Dempster, 1991; Rockstroh & Schweizer, 2001). It seems that for every reported correlation between each posited executive function and intelligence, counter evidence exists suggesting no such relationship.

In a review of the literature on the relationship between intelligence and resistance to interference, Dempster and Corkill (1999) concluded that it would be irresponsible to speculate the reasons for the contradictory findings: "The data are simply puzzling and that should be the take-home message" (p. 413). But the cause of such discrepancy between findings, like any puzzling result, is worthy of scientific investigation. Perhaps some other factor affects the relationship between fluid intelligence and the function in question, or perhaps an interaction between the function and some other component(s)

[2] Friedman et al. (2006) observed a significant correlation between fluid intelligence and antisaccade (oculomotor inhibition) performance (r = .23) but not between fluid intelligence and either stop-signal (behavioural inhibition) or Stroop (resistance to interference) performance (both r = .03).

of working memory is of critical importance to intelligence. More recent studies suggest that a relationship exists between maintaining task relevant information and resisting interference/inhibiting inappropriate responses. In another review paper, this time on the relationship between working memory capacity and fluid intelligence, Conway et al. (2003) concluded that working memory capacity reflects the active maintenance of goal-relevant information in the face of salient interference (such as proactive interference, response inhibition, or inhibition of a habitual but inappropriate response) and accounts for one-third to one-half of the variance in *g*. Indeed, other researchers stress the overarching importance of storage (relative to processing) aspects of working memory to intelligence.

Working memory storage and *g*

Recent evidence, from studies directly addressing storage versus processing accounts of the working memory capacity–*g* relationship, supports the central role of storage factors in mediating the relationship. In Chuderski et al. (2012), not only did their storage latent factor account for a much larger proportion of the variance in fluid intelligence than their attention control latent factor (70% relative to 25%), but attention control lost its ability to predict fluid intelligence when storage was accounted for. Similarly, Colom, Abad, Quiroga, Shih, and Flores-Medoza (2008) not only found that simple short-term storage (i.e., memory for numbers, letters, or a visual array) mediated the relationship between working memory and intelligence over and above processing speed, updating, and the control of attention, but also found that the relationship between these functions and both working memory capacity and intelligence disappeared when short-term storage was controlled for.

Indeed, a wealth of evidence shows that storage capacity is related to fluid intelligence in tasks that contain no obvious processing component (e.g., Chuderski et al., 2012; Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Colom et al., 2008; Cowan et al., 2005; Fukuda, Vogal, Mayr, & Awh, 2010). Retrieval from secondary

memory (i.e., long-term memory), such as recalling paired associates, is also related to Raven's performance (Mogle, Lovett, Stawski, & Sliwinski, 2008). In a similar vein to the literature on the relationship between fluid intelligence and other cognitive functions, however, counter evidence exists. Unsworth and Spillers (2010), for example, found that working memory capacity continued to predict Raven's performance when retrieval from secondary memory was accounted for.

Fukuda et al. (2010) report observations on the relationship between fluid intelligence and different aspects of working memory storage demand. In their visual change detection task, an array of four to eight geometric shapes (e.g., rectangles or ovals containing crosses or lines) could change in either a small way (e.g., a rectangle containing a line changing to a rectangle containing a cross) or a large way (e.g., a rectangle containing a line changing to an oval containing two lines). They found evidence for two distinct aspects of working memory storage, the number of stored representations and the resolution or complexity of stored representations, which respectively determine the abilities to detect large, and small, changes in the visual arrays. Interestingly, only the number (r = .66), and not the resolution (r = -.05), of items in working memory was significantly related to fluid intelligence.

In sum, many authors argue that equating fluid intelligence with working memory is unfruitful because it replaces one poorly-defined concept with another. One cannot deny, however, the sheer abundance of evidence linking fluid intelligence with working memory. The pressing unresolved issues are:

- Which specific aspect of working memory (e.g., storage vs. processing; a general processing function vs. a combination of different working memory functions vs. a specific executive function such as inhibition) mediates the working memory–*g* relationship?
- If a combination of different working memory functions work together to impact on intelligence test performance, what is the weighting of each of those functions?

- Are the relationships between fluid intelligence and storage/processing components of working memory influenced by another function necessary for the representation of complex tasks, such as task modelling?

Working memory for task rules and *g*

Task modelling

An intriguing new contender for contributing to the recruitment of Spearman's *g* is the "*task model* – a working memory description of the relevant facts, rules, and requirements used to control current behaviour" (Duncan et al., 2008, p. 140). In a number of studies, Duncan and colleagues have explored the strong relationship between Spearman's *g*, task modelling, and goal neglect (i.e., neglecting some task goal or requirement despite a preserved knowledge of that requirement). They argue that task modelling may provide a cognitive basis for *g* that is capable of explaining the occurrence of positive manifold among test scores.

Goal neglect was first observed by Duncan, Emslie, Williams, Johnson, and Freer (1996) in a letter monitoring task. Participants viewed letter pairs (interspersed with number pairs) and were required to read aloud the letters on one side of the display only; the appropriate side was initially indicated by a cue (WATCH LEFT or WATCH RIGHT), and then some way into the task, a prompt (– for left or + for right) indicated the side to read from for the remainder of trails. Duncan et al. noticed that some participants neglected the prompt (and continued to read on the side indicated by the initial cue) despite accurately recalling task instructions post task execution and being capable of responding accurately when given error feedback. Goal neglect was more common in low scorers (relative to high scorers) on a standard test of *g*—Cattell's Culture Fair test (Cattell, 1971; Cattell & Cattell, 1973)—and in people with restricted frontal lesions suggesting that goal neglect is associated with intelligence and frontal function. This study shows that the relationship between working memory and *g* cannot be explained on the basis of a straightforward storage function because their participants

were capable of describing the requirements of the task but some of them had difficulty responding to those requirements during task performance.

Considering the factors associated with goal neglect and the relationship between goal neglect and *g*, Duncan et al. (2008) dissociated complexity in real-time task execution (processing demand) from complexity in task instructions (storage demand for task requirements). To assess the effects of real-time execution demand, the letter monitoring task (of Duncan et al., 1996) was manipulated in two ways. First, participants viewed two (low-complexity) or four (high-complexity) characters (letters and digits[3]) per trial and were required to read the letter on the appropriate side. Second, participants viewed pure (low-complexity) or mixed (high-complexity) blocks of letter and digit pairs and were required to read the letters or add the numbers on the appropriate side. To assess the effects of task instruction complexity, participants were given *full instructions* (high-complexity) or *reduced instructions* (low-complexity), but the level of real-time demand was held constant (i.e., all participants completed one pure letter block and one pure digit block). In full instructions, participants were given instructions for both tasks prior to the first block (and were told that one task could be temporarily discarded); in reduced instructions, participants were not given instructions for the second block until the first block was complete. Neither manipulation of real-time task complexity affected the frequency of goal neglect, but goal neglect was more common in the full, relative to reduced, task model condition.

Finally, Duncan et al. (2008) highlighted how task model complexity is related to Spearman's *g* in a modified feature match task (Bright, 1998[4]) demanding inhibition of a prepotent but inappropriate response tendency. Participants viewed rapidly presented pairs of digits (numbers 1, 2, and 3) which were surrounded by coloured shapes (squares, circles, and triangles coloured in red, blue, or green). Single-matching pairs (i.e., pairs that matched in either colour or shape) required a response on the side with

[3] Note that the two-character version comprised six digit frames and seven letter frames, and the four-character version comprised six digit frames and seven letter-plus-digit frames (with one letter and one digit on each side).

[4] Experiments 1, 2, 3, and 6, presented in this thesis, employ this paradigm.

the largest number, but non-matching pairs (i.e., pairs that matched in neither colour nor shape) and double-matching pairs (i.e., pairs that matched in colour and shape) required no response. The low-complexity group received instructions for (and practiced) surround trials only, whereas the high-complexity group also received instructions for no-surround trials (which was to state the sum of the pair of digits), but these trials were never actually presented. Duncan et al. viewed the task as having three rules (corresponding to non-, single-, and double-matching pairs) and observed that, typically, more rules were failed in the full, but not the reduced, instructions condition; that is, two or more rules were failed in 68% of full-instructions participants relative to only 28% of reduced-instructions participants. Furthermore, the correlation between the number of rule failures and Culture Fair performance was numerically greater for full-instructions participants (r = .55) relative to reduced-instructions participants (r = .30), although the size of these coefficients was not statistically compared. Their findings indicate that the efficiency with which a task is cognitively modelled, or held in mind, may be of crucial importance to the recruitment of *g*.

Dumontheil, Thompson, and Duncan (2010) strengthened these conclusions using fMRI techniques. Participants completed eight variations of a task (i.e., words, shapes, letters, faces, dice, playing cards, animals, and arrows) in which A trials (A1 and A2) required one of two key press responses depending on the occurrence of specific stimuli, and B trials (B1, B2, and B3) required either a key press response or no response at all depending on the nature of the stimuli. For example, in the shapes task, A1 keypress responses were required for hollow shapes whereas A2 keypress responses were required for no shapes, and a response was required (on the side containing a greater number of dots) for single-matching pairs whereas no response was required for non- and double-matching pairs. In full instructions, participants were given instructions for each type of trial and were then told that only B trials, or a mixture of A trials and B trials, would be presented. In reduced instructions, participants were given

instructions for, and completed, B trials only. The authors found that B trial error was greater in the full, relative to the reduced, instructions group.

The existence of a relationship between task modelling, goal neglect, and Spearman's *g* suggests that *g* (as well as the working memory capacity–*g* relationship) cannot be explained on the basis of straightforward processing or storage functions. Indeed, Duncan, Schramm, Thompson, and Dumontheil (2012) found that the correlation with fluid intelligence is stronger for working memory for task rules than for more traditional measures of working memory including complex span tasks and visual short-term memory. Perhaps, then, this task modelling function is key to understanding individual differences in *g*.

The Multiple Demand Network

The *multiple demand network* (Duncan, 2010a), a pattern of fronto-parietal brain activation that is associated with the performance of a wide range of tasks, has been linked with task modelling and *g*. Others have termed these activation patterns as a *task positive network* (Fox et al., 2005), *task-activation ensemble* (Seelay, et al., 2007) or *task control network* (Dosenbach et al., 2006). The multiple demand network has been found to support performance of standard tests of fluid intelligence (e.g., Bishop, Fossella, Croucher, & Duncan, 2008), and damage to regions inside the network predicts loss in fluid intelligence but damage to regions outside the network does not (Woolgar et al., 2010).

Single cell studies of the primate cortex (e.g., Kusunoki, Sigala, Gaffan, & Duncan, 2009; Sigala, Kusunoki, Nimmo-Smith, Gaffan, & Duncan, 2008) support a network that codes information relevant to the current situation. These studies involve training monkeys to fixate their gaze to the centre of a screen and, when they detect a target picture in a series of pictures presented to the left or right of the screen at random, they are required to direct their gaze to the target. Targets can remain fixed throughout training or can be indicated by a cue at the beginning of each trial (there are usually

three cues which correspond to three different targets). Sigala et al. (2008) showed how prefrontal cells code the successive steps required for successful task completion by calculating the mean firing rates of a randomly selected sample of 324 lateral prefrontal neurons at three separate task events (cue, delay, and target). Each task event was characterised by a unique pattern of activity, suggesting that complex behaviour demands transition between one pattern of activity and another. These studies have shown that at least 50% of neurons in the prefrontal cortex discriminate non-targets from targets (both fixed and cued) and thus demonstrate selective coding of task-relevant information.

Indeed, primate neurons selectively code cognitive context (e.g., the identity and location of task-relevant objects, features, and categories) in prefrontal neurons both across (Freedman, Riesenhuber, Poggio, & Miller, 2001) and even within (Rao Rainer, & Miller, 1997) trials. Lesion studies support these findings showing that when monkey prefrontal and temporal cortices are disconnected, deficits are predicted by a requirement to integrate visual information across successive task parts (e.g., Browning & Gaffan, 2008). Furthermore, fMRI studies on humans using adaptation or multivoxel pattern analysis[5] also suggest selective coding of task-relevant information across frontal and parietal regions reflecting changes to an attended object (Hon, Epstein, Owen, & Duncan, 2006) or feature (Thompson & Duncan, 2009). Woolgar, Thompson, Borr, & Duncan (2011) found that fronto-parietal coding for stimulus–response mapping rules was stronger than that for specific instances of responses.

Duncan (2010b) clearly links the multiple demand network with both task modelling and fluid intelligence. He suggests that when a novel task is performed, the multiple demand network identifies and separates different aspects of the task forming a model of current behavioural goals. Such mental programming involves a series of

[5] Multivoxel pattern analysis (Haxby et al., 2001) involves applying pattern-classification algorithms to multiple voxels (rather than on single voxels as in traditional fMRI techniques) to decode the patterns of activity. Multivoxel pattern analysis has two main advantages over the use of traditional fMRI methods: (a) greater sensitivity, rendering it easier to find significant brain activation; and (b) reduced noise.

separate *attentional episodes* relating to specific instances in task performance. Each attentional episode involves only information and operations relevant to the current part of the task; when the current part of the task is complete, the attentional episode is succeeded by a new attentional episode reflecting the new demand of the task. Indeed, Dumontheil et al. (2010) found that the presentation of each task rule was associated with activity in parts of the multiple demand network, which rapidly returned to baseline in the 10–20 s delay between the presentation of each rule. This baseline activity was greater when rules were presented with other rules (full-instructions) relative to when the very same rules were presented at separate points in time (reduced-instructions). They argue that the smaller multiple demand activity changes that were associated with the presentation of later rules reflected weaker representations; this may explain why goal neglect is more frequent on later-specified task constraints. In sum, multiple demand regions "show neural properties suitable for the needs of sequential, multi-step behaviour, with selective focus on many kinds of task-relevant information" (Duncan, 2010a, pp. 176–177) and may provide a clearly defined neural basis for task modelling (and *g*).

The Global Workspace Theory

The *global workspace theory* (Dehaene, Kerszberg, & Changeux, 1998) may also be useful in understanding the biological basis for the task model. The global workspace represents a network of long-range excitatory axons (originating predominantly from the dorsolateral prefrontal cortex and inferior parietal cortex) that interconnect multiple distributed modalities and that is associated with a subjective feeling of conscious effort. Routine tasks, conversely, are associated with activation of specific low-to-medium range connections within parallel and distributed functionally specialised processors ranging from primary sensory (which receive sensory information from thalamic nerve projections; e.g., area V1) and unimodal processors (which combine multiple inputs within a specific modality; e.g., area V4) to heteromodal processors

(which "extract" highly processed categorical and semantic information; e.g., mirror neurons [area F5] or visuo-tactile neurons [lateral interparietal cortex]).

Dehaene et al. (1998) subjected a computational model of the global workspace theory to variants of the Stroop. In *effortless* versions of the Stroop (e.g., naming the ink colour of congruent colour words) the model responded correctly without activating the workspace, relying only on processors (e.g., direct one-to-one connections from colour-to-name units). In *effortful* versions, however, the model required global workspace activation to perform correctly (until the task was routinised). For example, when performing the effortful task of naming the ink colour of incongruent colour words, the model initially recruited processors which applied the inappropriate but prepotent word naming task to performance resulting in a series of errors. The negative rewards signalled by these errors increased vigilance; this *search phase* was associated with variable patterns of workspace activation reflecting the exploration and evaluation of various response rules. In the *effortful task execution phase* (after around 30 trials), the workspace activation stabilised resulting in correct performance.

Workspace theory suggests that as a task is practiced, correct performance does not require workspace activation and can instead rely on processors. Indeed, after the model produced a few correct responses whilst recruiting the workspace, vigilance and workspace activation was reduced resulting in immediate error that was followed by reactivation of the workspace. However, with further practice, performance can rely solely on processors because the Hebbian rule that is applied to processor units decreases word-to-name connections and increases colour-to-name connections. Five major processors contribute to the workspace: (a) *perceptual circuits* represent information regarding the present state of the external world (such as an object or discourse); (b) *motor programming circuits* allow the contents of the workspace to guide intentional behaviour (from low-level actions and gestures to high-level abstract plans); (c) *long-term memory circuits* represent information regarding past percepts (mental impressions of something perceived by the senses) and events; (d) *evaluation*

circuits assign a positive or negative value to workspace representations and maintain or alter the network depending on the value; and (e) *attention circuits* amplify or attenuate signals from processor neurons and update workspace representations independently from the external world. High-level motor and language circuits allow the active workspace representation to be described using gestures and words (Weiskrantz, 1997).

Supporting the interconnection of the workspace to high-level specialised modalities (i.e., perceptual, motor, memory, attentional, and evaluation processors), Goldman-Rakic (1988) identified a dense network of connections linking the dorsolateral prefrontal and inferior parietal cortices to the anterior and posterior cingulate, temporal, and parahippocampal cortices when monkeys completed an oculomotor delayed response task. The hippocampus may be involved in detecting novelty (Gray, 1994). Indeed, Deheane et al. speculate that the workspace may also involve a novelty detection mechanism and self-representations (allowing us to reflect on our own internal processes), and admit that workspace neurons may not be "functionally equivalent but rather may be organized in multiple hierarchically nested specialized circuits" (p. 14,534). Nonetheless the workspace theory, when viewed against research on the multiple demand network, may help to inform understanding of the biological and functional aspects of task modelling in the control of attention.

Research Framework

Chapter 1 has highlighted that the cognitive basis of *g* is still very much debated. Statistically, *g* accounts for the observation of positive manifold among test scores, so it is not surprising that so many different aspects of cognition have been related to *g* in the literature. An emerging consensus links *g* very closely to working memory, but the specific aspect of working memory that is most important to *g* is still disputed, with researchers debating the overarching importance of executive attention, inhibition, memory maintenance and retrieval, and task modelling in mediating the working memory–*g* relationship.

Although a complete understanding of the nature of *g* is beyond the scope of this thesis, the topics presented in Chapter 1 invite several avenues of investigation. The chosen approach was to determine which aspect of working memory is fundamental to the recruitment of *g* in task performance by manipulating both the complexity of the task model and demand for other postulated "risk factors" for the involvement of *g* (inhibition, updating, maintenance, and task switching). Such a systematic investigation of working memory processing and storage versus task modelling to the recruitment of *g* in task performance has not yet been undertaken. A secondary purpose of these investigations was to explore the factors associated with task model complexity and the breakdown of goal-directed behaviour.

Manipulations of real-time processing demand in task performance involved adapting two tasks that have been published in studies presenting findings on the cognitive correlates of *g*. Experiments 1 through 3 employed Bright's (1998) "Cognitive Reversal" feature match task; this task manipulated demand on inhibition because a requirement to inhibit a prepotent response tendency was present in some trials but was absent from others. Experiments 4 and 5 employed MacDonald et al.'s (2005) Dot Pattern Expectancy (DPX) task; this task was modified to separate and manipulate demand on a number of cognitive functions (working memory updating, working memory maintenance, inhibition of a prepared response, inhibition of a prepotent response) such that demand was high or present on some trials and low or absent on other trials. Experiment 6 involved a task comprising a mixture of cognitive reversal trials and DPX trials in an attempt to manipulate demand on task switching.

Manipulations of task modelling demand involved presenting task instructions for each of the experimental tasks in varying formats. Whilst the amount of executable task-relevant information presented to participants was held constant, the number of rules in which this information was presented was manipulated; instructions were presented as two distinct rules or four distinct rules (with the four rules essentially reflecting each of the two rules of two-rule instructions split into two chunks each).

These manipulations of task model complexity differ to the "full" and "reduced" distinction made by Duncan and colleagues which involved providing irrelevant information in one set of instructions and not in another. Another novel investigation in this thesis concerned the process (and relation to *g*) of re-modelling the task, that is, being required to or choosing to alter/update the information in the task model.

CHAPTER 2

Response Inhibition and Spearman's *g*

Overview

Chapter 2 explores the relationship between Spearman's *g* and response inhibition in a task for which instructions are presented as two distinct task rules. In Experiment 1, participants completed a computer-based speeded response task which required the inhibition of a prepotent yet inappropriate response to specific stimuli; these *critical items* were present in one block of trials and absent from another. It was assumed that presenting task requirements to participants as two separate rules would encourage the requirements to be conceptualised in the participants' minds in an efficient way (relative to, for example, presenting four separate rules). The errors produced by participants predominantly reflected a failure to withhold a prepotent response to critical items, and a greater amount of overall error was observed in the block that contained these items relative to the block that did not. However, performance on critical items was not significantly correlated with *g*, and the strength of correlations between performance and *g* was statistically equivalent across blocks. These findings indicate that response inhibition may not be a primary risk factor for the engagement *g* in the context of an explicitly defined two-rule task model.

Experiment 1

A strong relationship between scores on tests with high *g* loadings and scores on tests of working memory capacity is well documented (e.g., Ackerman et al., 2002; Conway et al., 2002; Süß et al., 2002; Unsworth et al., 2010). Researchers tend to agree that performance on complex working memory span tasks, such as the Reading Span (Daneman & Carpenter, 1980), relies on basic short-term memory processes like encoding, maintenance, and retrieval as well as controlled attention or executive

processes (for, respectively, remembering the last word from each read sentence, and preventing the task of reading the sentences from interfering with remembering the words; e.g., Engle et al., 1999). However, the relative roles of information processing versus storage in mediating the relationship between *g* and working memory are still debated. Some authors argue for the central role of some executive function or attentional mechanism in driving the relationship (e.g., Conway et al., 2003; Engle, 2002; Engle et al., 1999). Others instead argue for the fundamental importance of short-term storage, owing to the observation that the correlation with *g* for attention control (Chuderski et al., 2012; and mental speed, Colom et al., 2008) disappears when storage variables (e.g., the immediate recall of digits, letters, or spatio-temporal patterns) are controlled for.

The ability to resist interference or salient but inappropriate responses is sometimes posited as the single most important factor in determining individual differences in working memory capacity (e.g., Hasher & Zacks, 1988; Lustig, Hasher, & May, 2001; May, Hasher, & Kane, 1999). From this view, individuals that perform well on tests of working memory capacity are able to restrict working memory access to task-relevant information, resolve response conflict, and inhibit dominant but inappropriate responses. Chapter 1 highlighted the contradictory evidence surrounding the relationship between psychometric intelligence and tasks that are assumed to require inhibition; some studies support the existence of strong correlations between fluid intelligence and inhibition measures (e.g., Polderman et al., 2009; Dempster et al., 1995; Salthouse et al., 2003) whereas others present evidence for little relation (e.g., Friedman et al., 2006; Jensen, 1965; Jensen & Rowher, 1966; Jensen, 1964).

Although the correlation between intelligence and inhibition may depend on how these constructs are defined and measured, the conflicting findings nonetheless indicate that the intelligence–inhibition relationship is complicated. Perhaps interaction with some other task component(s) is of critical importance in driving the relationship between response inhibition and *g*. For example, findings by Duncan and colleagues

(e.g., Duncan et al., 2008; Duncan et al., 2012) suggest that the efficiency with which task requirements are conceptualised in mind (in what they term a task model) may be more fundamental to the task-involvement of Spearman's *g* than the real-time processing demands associated with the task. Across a series of computer based experiments they showed that the form in which instructions were presented to participants was the primary factor predicting both the level of goal neglect and the size of correlation between goal neglect and *g*. Thus, although increased real-time task complexity did not increase the level of neglect of task demands, an additional "dummy" requirement, which had no impact on what participants were required to do during actual task execution, increased the level of neglect and the strength of the performance–*g* correlation.

The main purpose of Experiment 1 was to establish whether and how the recruitment of *g* is affected by the presence of a requirement to overturn salient responses in the context of a low-complexity task model or *task conceptualisation*. The experimental task adopted was based on a feature match task (designed by Bright, 1998) which places demand on the inhibition of a prepotent response tendency. Previous studies employing modified versions of this task have reported significant correlations with *g* (e.g., Bright, 1998; Duncan et al., 2008). In the version of the task used in Experiment 1, pairs of coloured shapes appeared on screen that matched in colour or shape (termed *single-matching items*), colour and shape (termed *double-matching, or critical, items*), or neither colour nor shape (termed *non-matching items*). Single-matching items required one of two button-press responses (the specific response depended on other stimulus characteristics), whereas non- and double-matching items required no response. A button-press response (rather than no response) became the dominant response to items that matched because single-matching items were presented more frequently (30% of trials) than double-matching items (7.50% of trials). Thus, correct performance of double-matching items required the inhibition of a prepotent but inappropriate go response. One block of trials contained these critical items and another

block did not. Thus, real-time demand for response inhibition was manipulated by both item type and block type.

The chosen method for imposing order on the way in which the task was conceptualised in the participants' minds differed from that employed by Duncan and colleagues. Here, information regarding task requirements was split into two chunks, forming two distinct task rules. This was assumed to encourage a task model that was low in complexity relative to, for example, a four-rule task model that may be arguably higher in complexity. A number of performance measures, including a measure reflecting performance on double-matching items, were correlated with scores on a well-known standard test of fluid intelligence. These performance–*g* correlations (as well as performance scores) were compared across different parts of the task in order to gain an understanding of how the recruitment of *g* is affected by the presence of response inhibition demand in the context of a two-rule task conceptualisation.

On the basis of published evidence for a link between inhibition and *g*, it was predicted that prepotent but inappropriate button-press responses to double-matching items (termed *critical errors*) would be significantly correlated with the number of errors produced on the Culture Fair (termed *Culture Fair error*). However, due to Duncan et al.'s (2008) observations that the recruitment of *g* in task performance is unaffected by the level of real-time processing demand inherent in a task, it was also predicted that the correlation between Culture Fair error and performance would be statistically equivalent across the block that contained critical items (termed the *inhibition block*) and the block that did not (termed the *non-inhibition block*).

Method

Participants

An opportunity sample of 50 adults (30 female) aged between 18 and 62 years (M = 30.98, SD = 12.56) was recruited from the Department of Psychology, Anglia Ruskin University and the wider community. None of the participants had a history of

neuropsychological disorder. Psychology students were recruited via an on-line recruitment system and received course credit for their participation; this credit enabled the students to use the recruitment system for their own research projects. Community volunteers did not receive any payment. The sample size was based on relevant published research of a correlational nature (e.g., Duncan et al., 2008). The number of Culture Fair errors produced by participants ranged between 2 and 18 (M = 9.98, SD = 4.09) out of a maximum of 46.

Materials

Test of 'g': Culture Fair. General intelligence, or *g*, was measured using Cattell's standardised Test of 'g': Culture Fair, Scale 2, Form A (Cattell, 1971; Cattell & Cattell, 1973; hereafter termed the Culture Fair), which loads highly onto *g* at r = .81. The Culture Fair comprises four sub-tests that use geometrical figures to measure novel problem-solving ability in a set amount of time: series completions (3 min), classifications (4 min), matrices (3 min), and topological relations (2½ min).

Colour shape match task. The colour shape match task was programmed in E-Prime (Schneider, Eschman, & Zuccolotto, 2002) and was administered on a Dell PC. Stimulus frames were pairs of coloured outline objects, one containing a tick in the centre and the other containing a cross, presented in the centre of a high resolution colour monitor. The background screen was white (RGB: 255, 255, 255). Objects varied along two dimensions, colour and shape, each having three possible values: red (RGB: 255, 0, 0), blue (0, 0, 255), and green (0, 255, 0) for colour; circle, square, and triangle for shape. Objects were 12.70 mm × 12.70 mm and the distance between objects in a pair was 7.40 mm. Stimulus frames were a mixture of non-matching pairs (objects of different colours and shapes), single-matching pairs (objects of the same colour *or* shape), and double-matching pairs (objects of the same colour *and* shape). The task comprised two blocks, the inhibition block and the non-inhibition block, each containing 12 trials of 10 frames. In each block there were three single-matching pairs

in every trial, which were always presented as two of the first five frames (with at least one non-matching pair in between) and the ninth or 10th frame. In the inhibition block only, a double-matching pair was presented as the seventh frame in 75% of trials. All remaining frames were non-matching. There were three sub-blocks in each block, each containing four trials. In each sub-block, one trial comprised only colour-matching single-matching pairs, another trial comprised only shape-matching single-matching pairs, and the other two trials comprised a mixture of colour- and shape-matching single-matching pairs. Additionally, one trial required only left responses, one trial required only right responses, and two trials required a mixture of left and right responses. A double-matching pair was present in three of the four trials in each sub-block in the inhibition block only. Figure 2.1 displays a typical inhibition block trial in the colour shape match task.

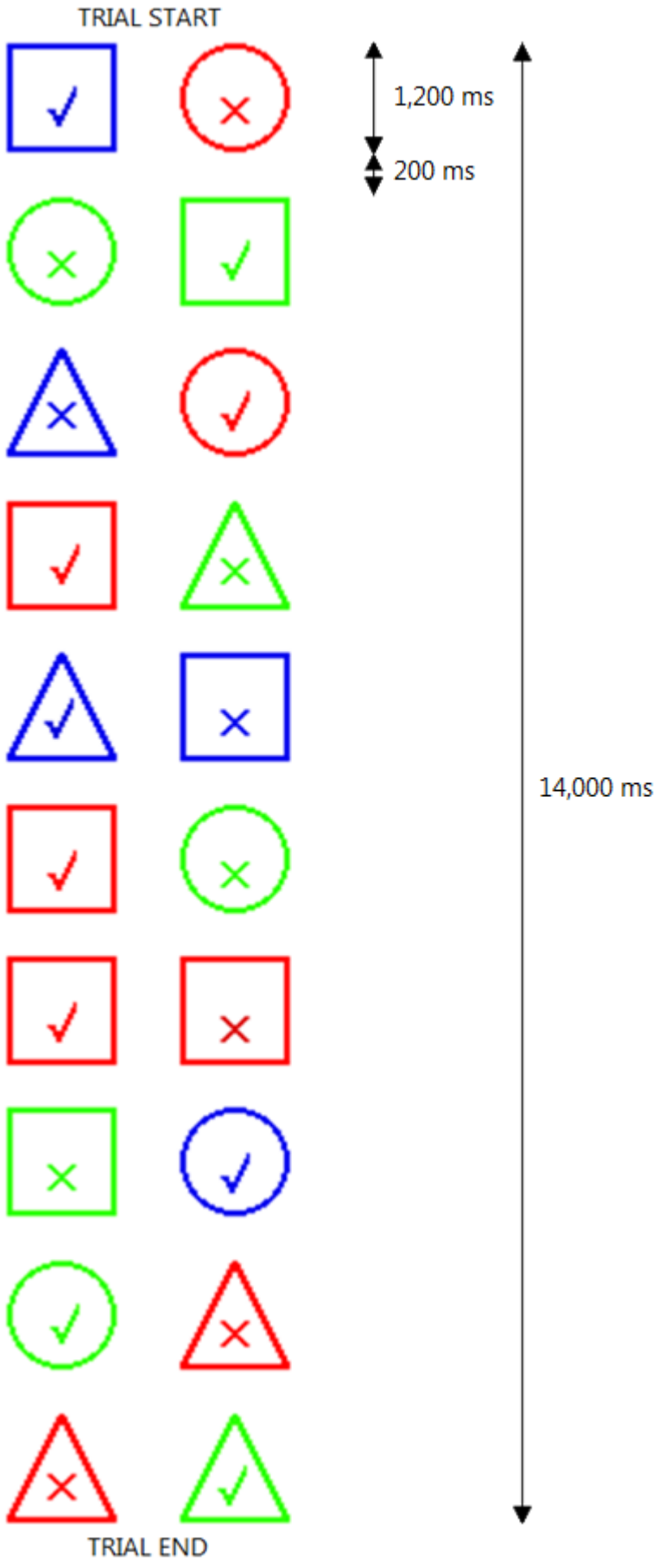


Figure 2.1. A typical trial in the colour shape match task.

Design

Performance measures for the colour shape match task are presented in Table 2.1. The criterion fail measure provided a criterion for success on a trial, taking into account performance across task elements; two criterion fail measures were calculated for the inhibition block (one that was determined, in part, by critical errors [termed *criterion fails*] and one that was not [termed *criterion fails without CEs*]) due to the fact that critical items were presented in this block only and were therefore not involved in the calculation of criterion fails in the non-inhibition block. A mixed design was adopted in which performance scores and correlations between performance scores and Culture Fair error were compared across performance measures (repeated measures), blocks (repeated measures: inhibition block vs. non-inhibition block), and order groups (independent samples: participants that completed the inhibition block first [termed the *inhibition–non-inhibition group, n* = 25] vs. participants that completed the non-inhibition block first [termed the *non-inhibition–inhibition group, n* = 25]). Other than the sequencing of blocks, the specific sequencing of trials was maintained for all participants in order to hold the level of prepotency associated with responses to critical items constant across participants.

Table 2.1

Definitions of performance measures for the colour shape match task

Performance measure	Definition
Critical error	Response to double-matching frame
Miss	No response to single-matching frame
Hand error	Response on incorrect side to single-matching frame
False positive	Response to non-matching frame
Response time	Time taken to respond to single-matching frame (ms)
Criterion fail (per trial)	Critical error
and/ or	Miss or hand error to first two single-matching frames
and/ or	Miss to final single-matching frame
and/ or	Three or more false positives
Criterion fail without CEs	Criterion fail excluding critical error as a criterion

Procedure

Participants were tested individually in a quiet and brightly lit testing room. The Culture Fair was completed first and was administered exactly as specified in the manual. The monitor was then placed approximately .50 m in front of the participant who sat on a height adjustable chair which allowed them to adjust the position of their head in relation to the monitor for optimum comfort.

Participants were informed that pairs of coloured objects, one containing a tick and the other containing a cross, would be presented one at a time in the centre of the screen. They were told that the objects may share either the same colour or the same shape, and that towards the end of each trial, they might see a pair that shares both the same colour and the same shape. Two distinct task rules were presented to participants. The first rule was "respond to items that match in colour *or* shape by pressing the side corresponding to the placement of the tick." The second rule was "ignore items that do not match in colour or shape, and items that match in both colour *and* shape." Participants worked through an example trial on paper (see Appendix B) and received feedback from the experimenter on their responses before hearing the rules for a second time. Before commencing the first block of trials participants were asked to verbally repeat the rules. If the rules were repeated incorrectly by the participant (i.e., with any omissions) the appropriate rule was presented again; the task did not begin until all rules were repeated correctly to prevent performance from reflecting poor understanding of task requirements. Verbal responses were recorded on a dictaphone. Participants were instructed to respond as quickly but as accurately as possible and, when they were ready, the experimenter pressed the mouse key to start the first block.

In each trial, after a blank-screen interval of 1,500 ms, 10 stimulus frames were presented. Each stimulus frame was presented for 1,200 ms with a 200 ms blank-screen interval in between each frame. Behavioural responses were made by pressing highlighted keys B and N (for left and right, respectively) on the computer keyboard and were recorded within E-Prime; a response was attributed to a frame if it occurred

within 200 ms (< 200 ms was considered anticipatory) and 1,200 ms (> 1,200 ms was considered an outlier) of stimulus onset. After each trial READY was presented on the screen until the experimenter pressed the mouse key to start the next trial. After the first block, participants were asked how many rules there were and to repeat the rules (if any of the rules were stated in a different format to that outlined in task instructions, or if there were any omissions, the appropriate rule was repeated and the participant was asked to state the rules again). Participants were reminded to respond as quickly but as accurately as possible and, when they were ready, the experimenter pressed the mouse key to start the second block. After the second block, participants were again asked how many rules there were and to repeat the rules. Each block took approximately 3 min to complete. See Appendix A for the full research protocol.

Results

Main findings

Performance. Performance scores are compared across blocks in Table 2.2. The data suggest that items requiring the inhibition of a prepotent response tendency were more difficult than items that contained no such requirement. Paired t-tests (two-tailed) confirmed that error was greater on the critical error measure relative to all other measures with the exception of criterion fails in the inhibition block (all $p < .001$, which was significant using the Bonferroni-corrected alpha level of $p < .05 / 9 = .006$). Table 2.2 shows that, across blocks, performance was worse in the inhibition block (relative to the non-inhibition block) for misses, response time, and criterion fails (all $p < .001$, which was significant using the Bonferroni-corrected alpha level of $p < .05 / 6 = .008$).

Table 2.2

Mean performance scores across blocks

	Block						
	Inhibition		Non-inhibition		Paired t-test		
Measure	*M*	*SD*	*M*	*SD*	*t*(49)	*p*	Cohen's *d*
Critical error	.40	.26	-	-	-	-	-
Hand error	.04	.04	.03	.04	.92	.36	.23
Miss	.20	.16	.13	.14	3.87	< .001	.53
False positive	.08	.19	.05	.17	1.59	.12	.22
Criterion fail	.50	.27	.18	.24	9.75	< .001	1.37
without CEs	.30	.28	.18	.24	4.17	< .001	.56
Response time	850	87	816	95	3.90	< .001	.57

Note. All means are expressed as proportion of total possible error for that measure with the exception of response time which is presented in ms.

***g* correlations.** Pearson's correlations (one-tailed) between performance scores and Culture Fair error (raw scores) are presented in Table 2.3. For a sample size of 50, the critical value for Pearson's *r* (one-tailed) is ±.24 at the $p < .05$ alpha level. The correlation between age and Culture Fair error was weak and non-significant[6], $r(47) = .16$, $p = .13$. The performance measures that were significantly correlated with Culture Fair error did not correspond to those measures that were more difficult (i.e., associated with more error). Culture Fair error was significantly correlated with hand errors, misses, and criterion fails, although the correlation was marginal for criterion fails without CEs, $r(48) = .22$, $p = .06$; significant correlations ranged between $r = .24$ and $r = .33$. Williams-Hotelling t-tests (two-tailed), which require a critical value of ±2.01 for significance at the $p < .05$ alpha level for a sample size of 50, confirmed that the correlation with Culture Fair error was (a) not significantly stronger for critical errors relative to other measures, and (b) statistically equivalent across blocks for every measure (all $p > .10$).

[6] This correlation was based on data from 49 (instead of 50) participants because age for one participant was not recorded.

Table 2.3

Pearson's correlations between Culture Fair error and performance across blocks

	Block					
	Inhibition		Non-inhibition		William's Hotelling t-test	
Measure	*r*(48)	*p*	*r*(48)	*p*	*t*(47)	*p*
Critical error	.20	.08	-	-	-	-
Hand error	.33	.01	.30	.02	.21	.84
Miss	.25	.04	.24	< .05	.08	.94
False positive	.14	.17	.19	.10	.47	.64
Criterion fail	.30	.02	.24	.05	.47	.64
without CEs	.22	.06	.24	.05	.17	.86
Response time	-.06	.34	.05	.37	1.15	.25

Culture Fair z-score charts. To clarify the relationship between each performance measure and Culture Fair error, participants were assigned a z-score bin (width of .50 SD) based on their Culture Fair raw error scores; Figures 2.2 and 2.3 display performance scores for each measure across Culture Fair z-score bins. The charts for hand errors, misses (both Figure 2.2), and criterion fails (Figure 2.3) reinforce the significant relationship with Culture Fair error observed for these measures. In the inhibition block, participants that scored ≥ 1 SD above the sample mean on the Culture Fair produced 14% fewer misses and 28% fewer criterion fails (relative to only 10% fewer misses and 19% fewer criterion fails in the non-inhibition block) than participants that scored ≥ 1 SD below the sample mean. Although the correlation between critical errors and Culture Fair error failed to reach significance, the chart for critical errors (Figure 2.2) demonstrates that 20% fewer critical errors were made by participants scoring ≥ 1 SD above, relative to participants scoring ≥ 1 SD below, the sample mean.

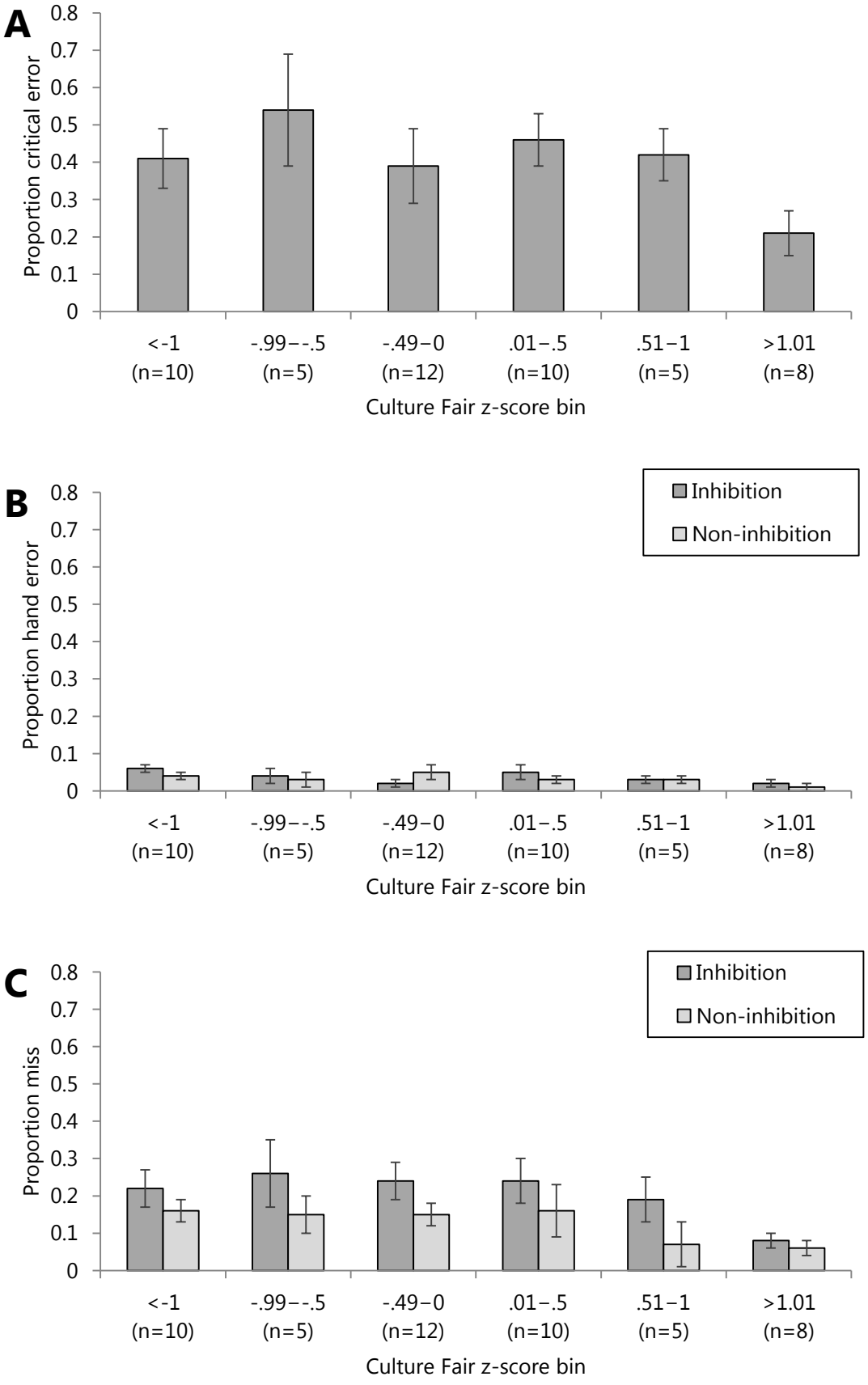


Figure 2.2. Mean performance (proportion error) across Culture Fair z-score bins for critical errors (A), hand errors (B), and misses (C). Error bars represent standard error.

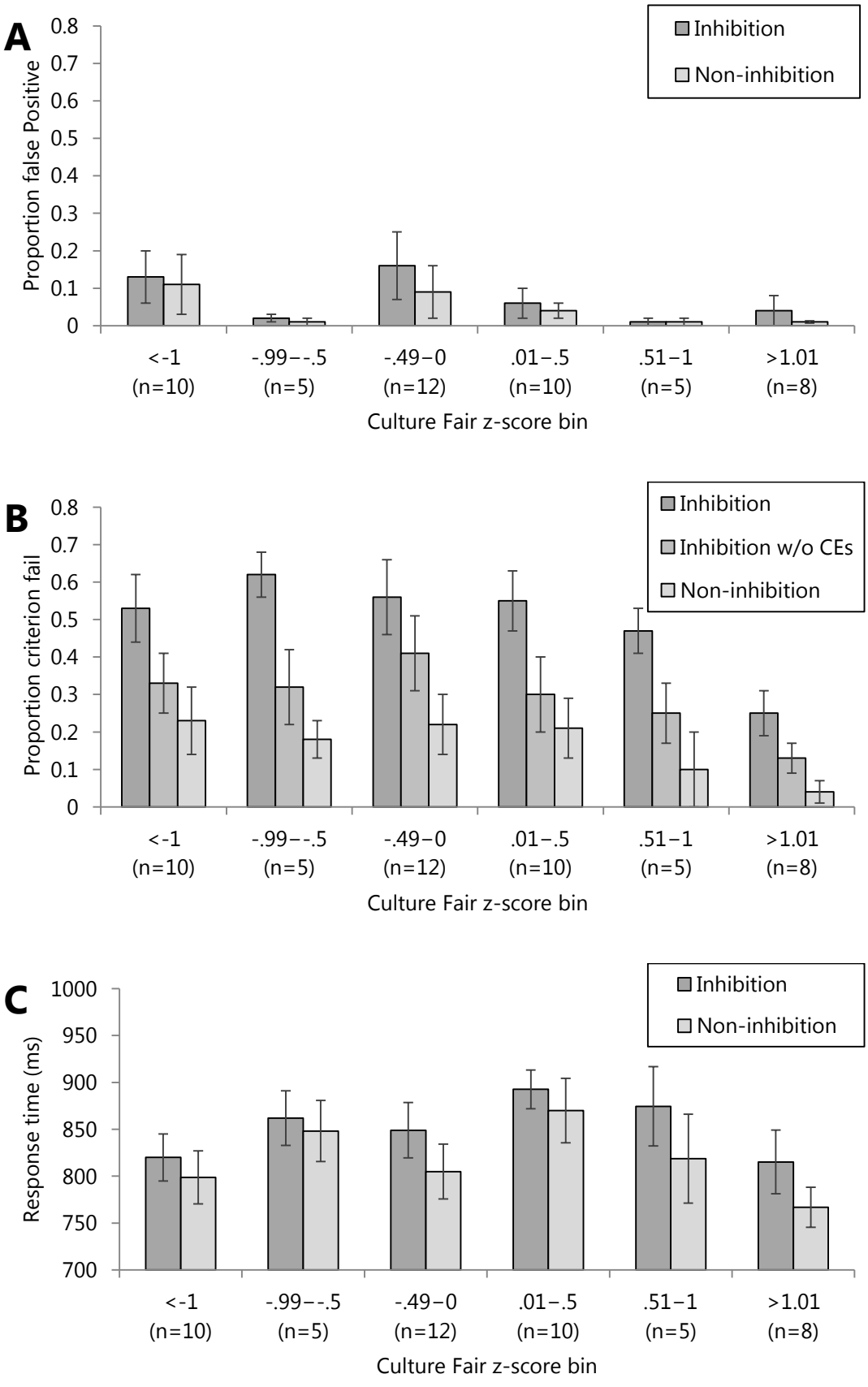


Figure 2.3. Mean performance across Culture Fair z-score bins for false positives (proportion error, A), criterion fails (proportion error, B), and response time (ms, C). Error bars represent standard error.

Additional findings

Practice effects. ***Performance.*** There was a general trend for reduced error and faster response time as participants progressed through the task. This is demonstrated in Table 2.4 which displays performance scores across the three sub-blocks in each block. Repeated measures analyses of variance (ANOVAs; two-tailed), which are presented in Table 2.5, revealed a significant improvement in performance (a) across sub-blocks 1 through 3 for hand errors, misses, and criterion fails in the inhibition block (although hand errors stayed constant after sub-block 2 due to ceiling effects), and (b) from sub-block 1 to sub-block 2 for misses, false positives, and criterion fails in the non-inhibition block (all $p < .05$). However, each sub-block comprised only four trials challenging the reliability of these comparisons.

Table 2.4

Mean performance scores in each sub-block

	Sub-block 1		Sub-block 2		Sub-block 3	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
			Inhibition block			
Critical error	.45	.28	.36	.37	.33	.33
Hand error	.06	.10	.02	.04	.02	.04
Miss	.24	.18	.19	.20	.17	.21
False positive	.10	.21	.09	.23	.06	.19
Criterion fail	.62	.29	.46	.36	.43	.29
without CEs	.38	.34	.29	.34	.25	.29
Response time	860	105	834	109	840	160
			Non-inhibition block			
Hand error	.04	.07	.03	.07	.02	.04
Miss	.22	.21	.09	.14	.09	.13
False positive	.10	.19	.06	.18	.06	.19
Criterion fail	.28	.32	.11	.25	.15	.26
Response time	818	157	815	106	800	100

Note. Means are expressed as proportion of error except for response time which is presented in ms.

Table 2.5

Repeated measures ANOVAs comparing mean performance scores across sub-blocks

Measure	*df*	*F*	*p*	η_p^2
		Inhibition block		
Critical error	2, 98	2.44	.09	.05
Hand error [a]	1.34, 65.61	6.14	.003	.11
Miss	2, 98	3.56	.03	.07
False positive [a]	1.54, 75.40	2.84	.06	.06
Criterion fail [a]	1.64, 80.50	13.49	< .001	.22
without CEs	2, 98	5.79	.004	.11
Response time [a]	1.28, 62.51	.89	.42	.02
		Non-inhibition block		
Hand error	2, 98	1.64	.20	.03
Miss [a]	1.63, 79.95	25.14	< .001	.34
False positive	2, 98	4.19	.02	.08
Criterion fail	2, 98	12.22	< .001	.20
Response time [a]	1.15, 56.48	.43	.65	.009

Note. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

Despite the occurrence of practice effects, satisfaction of the requirement to inhibit a prepotent response to critical items did not ensure later successful inhibition. Table 2.6 presents critical errors across trials (in the inhibition block) for three participants with differing Culture Fair scores[7], alongside the percentage of participants committing critical error on each trial. The percentage of participants that produced a critical error decreased as the task progressed (from 32% in the first trial to 15% in the final trial), yet successful inhibition of a prepotent response to critical items did not prevent later critical error for 78% of participants (including the three participants in Table 2.6). Typically, participants reacted immediately upon producing a critical error demonstrating that the failure to conform to task instructions was not due to straightforward forgetting of rules.

[7] These specific participants were chosen because their scores corresponded to relatively low, relatively medium, or relatively high scores on the Culture Fair.

Table 2.6

Pattern of critical error across trials for three participants with differing Culture Fair scores and percentage of participants committing critical error on each trial

Trial	Pattern of critical error			% of participants
	Participant A (2 Culture Fair errors)	Participant B (8 Culture Fair errors)	Participant C (16 Culture Fair errors)	
1	0	1	1	32
2	0	0	0	22
3	0	0	1	20
5	0	1	1	22
7	1	0	0	17
8	0	0	0	15
10	0	0	1	22
11	0	0	0	12
12	0	1	1	15
Overall	1	3	5	78

g correlations. Pearson's correlations (one-tailed) between sub-block performance scores and Culture Fair error are presented in Table 2.7. Consistent with the main findings correlations were, at most, moderate in strength. In the inhibition block, Culture Fair error was significantly correlated with critical errors in sub-block 1 only, with criterion fails in sub-blocks 1 and 3, and with hand errors and misses in sub-block 2 only. In the non-inhibition block, Culture Fair error was significantly correlated with hand errors, misses, and criterion fails, but only in sub-block 1 (as well as sub-block 3 for criterion fails only). Williams-Hotelling t-test (two-tailed) revealed that the correlation with Culture Fair error for criterion fails in the inhibition block was significantly stronger in sub-block 1 relative to sub-block 2, $t(47) = 2.15$, $p = .04$. All other comparisons were non-significant (all $p > .10$), but two marginal effects were observed in the inhibition block. The correlation with Culture Fair for critical errors was marginally stronger in sub-block 3 relative to sub-block 2, $t(47) = 1.78$, $p = .08$; the correlation with Culture Fair error for misses was marginally stronger in sub-block 2 relative to sub-block 1, $t(47) = 1.88$, $p = .07$.

Table 2.7

Pearson's correlations between Culture Fair error and sub-block performance

	Sub-block 1		Sub-block 2		Sub-block 3	
Measure	*r*(48)	*p*	*r*(48)	*p*	*r*(48)	*p*
			Inhibition block			
Critical error	.21	.07	-.04	.41	.26	.04
Hand error	.19	.09	.31	.02	.20	.08
Miss	.07	.32	.32	.01	.20	.08
False positive	.07	.31	.17	.12	.15	.15
Criterion fail	.38	.003	.15	.15	.26	.04
without CEs	.20	.08	.21	.07	.18	.10
Response time	-.11	.23	.01	.46	-.04	.39
			Non-inhibition block			
Hand error	.31	.01	.17	.12	.08	.29
Miss	.28	.03	.21	.07	.11	.22
False positive	.25	.04	.13	.19	.20	.08
Criterion fail	.25	.04	.11	.23	.24	.05
Response time	.16	.13	.06	.33	-.04	.39

Order effects. ***Performance.*** Performance scores are compared across order groups (inhibition–non-inhibition group vs. non-inhibition–inhibition group) in Table 2.8. There was a pattern of greater error in the block that was completed first. That is, inhibition block error appeared greater in the inhibition–non-inhibition group (relative to the non-inhibition–inhibition group), whereas non-inhibition block error appeared greater in the non-inhibition–inhibition group. However, independent samples t-tests (two-tailed) showed that the only significant difference in performance across order groups was for criterion fails in the non-inhibition block which were more frequent in the non-inhibition–inhibition group ($p = .04$). Culture Fair error was also statistically equivalent across order groups suggesting that differences in performance across order groups were not due to differences in *g*.

Table 2.8

Mean performance scores across order groups

	Order group				Independent t-test			Cohen's *d*
	Inhibition–non-inhibition		Non-inhibition–inhibition					
Measure	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	
Culture Fair	9.20	3.74	10.76	4.34	48	1.36	.18	-.39
			Inhibition block					
Critical error	.44	.28	.35	.25	48	1.19	.24	.34
Hand error	.05	.05	.03	.03	48	1.16	.25	.50
Miss	.23	.15	.18	.17	48	1.04	.30	.31
False positive	.09	.20	.07	.20	48	.39	.70	.10
Criterion fail	.55	.25	.44	.28	48	1.46	.15	.42
without CEs	.34	.28	.27	.27	48	.90	.37	.26
Response time	857	83	844	92	48	.53	.60	.15
			Non-inhibition block					
Hand error	.03	.03	.04	.05	48	-1.16	.25	-.25
Miss	.11	.16	.15	.11	48	-1.01	.32	-.30
False positive [a]	.01	.05	.09	.23	25.93	-1.61	.12	-.57
Criterion fail	.11	.18	.24	.27	48	-2.11	.04	-.58
Response time	825	94	807	97	48	.67	.50	.19

Note. All means are expressed as proportion of error except for Culture Fair error (number of errors) and response time (ms). $n = 25$ in each order group. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

g correlations. Pearson's correlations (one-tailed) between task performance and Culture Fair error are compared across order groups in Table 2.9. Correlations with Culture Fair error appeared noticeably larger in the inhibition–non-inhibition group (relative to the non-inhibition group) for misses, criterion fails, and response time (in both blocks). The opposite pattern of larger correlations with Culture Fair error in the non-inhibition–inhibition group was observed for false positives (in both blocks), hand errors (in the inhibition block) and critical errors. Fisher's z-tests (two-tailed), however, showed that none of the correlations differed statistically across order groups ($p > .10$).

Table 2.9

Pearson's correlations between Culture Fair error and order group performance

	Order group			
	Inhibition–non-inhibition		Non-inhibition–inhibition	
Measure	*r*(23)	*p*	*r*(23)	*p*
		Inhibition block		
Critical error	.19	.18	.29	.08
Hand error	.33	.05	.45	.01
Miss	.43	.02	.17	.20
False positive	.10	.32	.21	.16
Criterion fail	.52	.004	.22	.15
without CE	.38	.03	.15	.24
Response time	.17	.21	-.22	.15
		Non-inhibition block		
Hand error	.26	.10	.28	.09
Miss	.26	.11	.20	.18
False positive	-.04	.43	.21	.16
Criterion fail	.25	.12	.16	.22
Response time	.27	.09	-.10	.31

Note. $n = 25$ in each group.

Learning and reconceptualisation of task rules. Participants were asked to repeat task rules at three points during the experiment: at task instructions, after the first block, and after the second block. At each time point, participants were categorised as: (a) being correct (if they stated all rules without any omissions, not necessarily word-for-word, in the first instance) or incorrect (if they omitted parts of rules in the first instance[8]) in their description of task rules; and (b) having reconceptualised (if they stated fewer or more than two distinct rules) or not reconceptualised (if they stated two distinct rules) task rules. An independent t-test (two-tailed) revealed that Culture Fair error was significantly lower in participants that stated the rules correctly ($n = 33$, $M = 8.52$ errors, $SD = 3.99$), relative to participants that stated the rules incorrectly ($n = 17$,

[8] When participants made omissions in their description of task rules, the appropriate rule was repeated by the experimenter until the participant successfully repeated all parts of all rules. This was to reduce the likelihood of further performance reflecting forgotten rules, rather than an inability to attend to those rules.

M = 12.82 errors, *SD* = 2.53), at the time of task instructions, $t(48) = 4.05$, $p < .001$, $d = -1.32$; a marginal effect in the same direction was observed after the first block, $t(48) = 1.84$, $p = .07$, $d = -.61$. All other comparisons were non-significant (all $p > .10$).

Rule failure scores and total proportion error scores. For a more straightforward comparison with Duncan et al. (2008), participants were assigned rule failure and proportion error scores for each block. Rule failure scores were calculated by adding together the proportion of failed responses to each item type (non-, single-, and double-matching) in each sub-block; an item was failed if proportion error was greater than .75. Given that there were three sub-blocks per block, the highest possible rule failure score was 9 (3 sub-blocks × 3 items) in the inhibition block and 6 (3 sub-blocks × 2 items) in the non-inhibition block. Proportion error scores were calculated as the unweighted mean of the proportion of failed responses to each item type.

Table 2.10 displays performance scores and correlations with Culture Fair error for rule failure and proportion error scores. Paired t-tests (two-tailed) revealed that rule failure scores were significantly greater in the inhibition block relative to the non-inhibition block, $t(49) = 2.89$, $p = .006$, $d = .41$. Similarly, proportion error scores were significantly greater in the inhibition block relative to the non-inhibition block, $t(49) = 7.42$, $p < .001$, $d = 1.03$. However, Williams-Hotelling t-tests (two-tailed) showed that the correlation with Culture Fair error for both rule failure and proportion error scores were statistically equivalent across blocks (both $p > .10$).

Table 2.10

Performance and Pearson's correlations with Culture Fair error for rule failure and proportion error scores

	Inhibition block				Non-inhibition block			
	Performance		*g* correlation		Performance		*g* correlation	
Score	*M*	*SD*	*r*(48)	*p*	*M*	*SD*	*r*(48)	*p*
Rule failure	.44	.68	.21	.07	.14	.57	.13	.18
Proportion error	.24	.13	.34	.01	.11	.14	.28	.02

Discussion

The results of Experiment 1 show that inhibiting a response to critical items (i.e., items that were associated with a prepotent yet inappropriate response tendency) was more difficult for participants than responding appropriately to other task rules. Proportion of error was significantly greater on critical items (critical error) compared to all other measures; the only exception was criterion fails but this is unsurprising given that criterion fails took into account performance across all task elements. This finding supports other research that has found that no-go trial performance is typically less accurate than go trial performance (e.g., Redick et al., 2011). Proportion of error was also significantly greater in the block that contained critical items compared to the block that did not for a number of task measures (misses and response time to single-matching items and criterion fails). It is unclear whether the inhibition block was associated with greater error due to the inclusion of response inhibition demand per se, or was rather a consequence of the existence of an additional active task requirement.

One of the observations made by Duncan et al. (1996) was that "with brief trials that are all broadly similar … a requirement that has once been satisfied is not subsequently neglected" (p. 294). However, this is not supported by the present findings. For 78% of participants, successful inhibition of a response to critical items did not prevent later critical error. Critical items were also exempt from the effects of practice; a significant improvement across sub-blocks was not found for critical items, but was found for other measures (misses and criterion fails in both blocks, hand errors in the inhibition block, and false positives in the non-inhibition block). Perhaps a requirement to inhibit a prepotent but inappropriate response is particularly resistant to improvement via practice, even in cases when this requirement has previously been satisfied.

Counter to predictions, critical error was not significantly correlated with *g* (as measured by performance on the Culture Fair), despite a greater level of observed error on this measure. The only measures that were significantly correlated with *g* were hand errors, misses, and criterion fails in both blocks. Statistical comparisons of the strength

of the correlation with Culture Fair error across measures did not provide evidence for a stronger correlation with *g* for critical errors relative to other measures. There was also very little numerical (and no statistical) difference between correlations with *g* for equivalent measures across blocks (e.g., misses in the inhibition block vs. misses in the non-inhibition block). This was, however, what was expected on the basis of the findings from Duncan et al. (2008) which showed that the level of goal neglect, and the relationship between goal neglect and *g*, was unaffected by manipulations of the level of real-time performance demand. These findings contradict the literature that suggests a link between response inhibition and intelligence (e.g., Polderman et al., 2009), but support other studies that suggest that response inhibition is not of fundamental importance to the engagement of *g* (e.g., Friedman et al., 2006). However, it is possible that the correlations observed here may be strengthened by the administration of a more complex set of task instructions (e.g., Duncan et al., 2008).

Indeed, the results of Experiment 1 provide some support for the notion that the ability to efficiently learn task rules is related to *g*. When describing task rules, participants that did not make any errors or omissions (on their first attempt at stating the rules) fell significantly higher on the *g* distribution than participants that had difficulty repeating task rules correctly. Culture Fair scores were also compared across participants that imposed their own order on the instruction rules—that is, stated fewer or more rules than initially outlined in their task instructions—relative to participants that did not. These comparisons were run on the assumption that if the ability to form an efficient task conceptualisation is linked to *g*, participants that reconceptualise the task (and as a result reduce the number of chunks in which task requirements are represented in mind) may correspond to those participants that have higher *g*. However, no difference in Culture Fair error was observed across participants that did, and participants that did not, reconceptualise task rules.

It is possible that the postulated ability to reconceptualise the task was not measured reliably in Experiment 1 because, arguably, the two presented rules already reflected a

relatively efficient representation of the task. As such, reconceptualisation of task rules may have been unnecessary. Furthermore, participants were simply asked to repeat the rules, rather than to state the rules as they were *represented in their own mind*. Thus, any participants that may have reconceptualised the presented task rules may not have demonstrated this in their verbalisation of the rules. Instead, some participants may have viewed this part of the experiment as some kind of memory test in which rules were to be repeated exactly as they were presented. That some participants may have reconceptualised task rules in mind is indeed a problem with the design; for these participants, the task model that was employed when performing the task may have differed from that which was encouraged by the formal task instructions. This complicates conclusions made on the assumption that the format of explicitly defined instructions reflects the model employed during task performance.

In conclusion, both the item (critical double-matching pairs) and the block (the block that contained critical pairs) that required the inhibition of a prepotent but inappropriate response tendency, was associated with greater error. However, performance on these critical items/this critical block was not more sensitive to variations in *g* (relative to the items/block that did not require inhibition), at least in the context of an efficient task conceptualisation (where efficiency is defined as the number of distinct chunks in which task constraints are presented). Perhaps when task rules are held in mind efficiently participants have sufficient resources available to cope with the demands of the task so the recruitment of *g* is unnecessary. In Experiment 2, instructions for the same task were presented as four rules to determine whether the relationship between response inhibition and *g* increases as a function of task model complexity.

CHAPTER 3

Response Inhibition, Task Conceptualisation, and Spearman's *g*

Overview

Chapter 3 explores the relationship between Spearman's *g* and response inhibition in the context of a four-rule task model. When task requirements were presented as two rules, in Experiment 1, manipulations of response inhibition demand did not affect task-sensitivity to *g*. The method of Experiment 2 was carried over from Experiment 1, other than the presentation of task requirements as four, instead of two, distinct rules. Participants also received the same amount of operative task-relevant information, when given task instructions, as those in the previous experiment. The recruitment of *g* was significantly higher in the four-rule group of Experiment 2 relative to the two-rule group of Experiment 1, but this finding was limited to conditions that demanded response inhibition (i.e., critical errors and overall error in the inhibition block only). This was despite statistically equivalent performance scores across the two experiments. Moreover, manipulations of the presence of response inhibition demand affected the correlation between overall block performance and *g* in the present experiment only. These findings suggest that, when a task demands inhibition, the recruitment of *g* is strengthened by increasing the number of chunks in which task requirements are represented in mind; the relationship between response inhibition and *g* may also be contingent upon such task model complexity.

Experiment 2

The findings of Experiment 1 showed that response inhibition was not significantly related to *g* in the context of an explicitly expressed two-rule task conceptualisation. As explained in Experiment 1, it is possible that some other aspect of the performance of complex tasks may drive the correlation between response inhibition and *g*. This view is encouraged by studies showing that the relationship between attention control (and

mental speed) and *g* is driven by basic short-term storage functions (Colom et al., 2008; Chuderski et al., 2012), and by the finding that the relationship between performance and *g* (and the level of goal neglect) is affected by the complexity of presented task instructions (e.g., Duncan et al., 2008; Duncan et al., 2012). Similarly, support for the overarching importance of storage rather than processing factors in *g* is supported by the finding that individual differences in working memory capacity are determined by the ability to maintain and retrieve information (rather than execution attention and inhibition; Redick et al., 2011).

The purpose of Experiment 2 was to consider whether increasing the complexity of the task model, relative to the model encouraged in Experiment 1, would enhance the recruitment of *g* in task performance. All aspects of the materials and procedure (including all stimuli, presentation rates, etc.) were identical to Experiment 1, with the exception of the format of presented task instructions. Task model complexity was manipulated in a different way to published studies (e.g., Duncan et al., 2008) in which one group of participants were given the instructions only for the task to be completed (reduced-instructions), and another group of participants were given further instructions for a task that was never actually performed (full-instructions). Here, the amount of operative task-relevant information presented to participants was instead held constant; the instructions of Experiment 2 were formed by splitting the information inherent in each of the two rules (of Experiment 1) into two parts each, forming four separate "chunks" of information. It was assumed that the required use of four separate task rules during task performance would be associated with more demand (i.e., would require more working memory or attentional resources) than the required use of two task rules.

In the same manner as in Experiment 1, a number of performance measures were correlated with scores on the Culture Fair. These performance–*g* correlations (as well as performance scores) were compared across different parts of the task, and were also compared to those observed in Experiment 1. This method enabled an exploration of

(a) whether and how the relationship between response inhibition and *g* is affected by the number of rules presented at task instructions, and (b) the extent to which the relationship between task conceptualisation and *g* (if any) is dependent on the requirement to inhibit a prepotent response tendency.

On the assumption that a larger (and perhaps less efficient) body of rules is likely to increase the recruitment of *g* in task performance, it was predicted that the correlation between performance and Culture Fair error would be: (a) significant for items requiring the inhibition of a prepotent but inappropriate button-press response (critical errors); (b) significantly stronger in the block that contained critical items (inhibition block) relative to the block that did not (non-inhibition block); and (c) significantly stronger in the four-rule group of Experiment 2 relative to the two-rule group of Experiment 1. Given the evidence for increased goal neglect when task instructions are more complex, it was also predicted that performance would be significantly poorer in the four-rule group relative to the two-rule group.

Method

Participants

An opportunity sample of 50 adults (33 female) aged between 18 and 63 years (M = 29.22, SD = 12.78) with no history of neuropsychological disorder was recruited from the Department of Psychology, Anglia Ruskin University (via an on-line recruitment system which granted students course credit) and the wider community (community volunteers did not receive any payment). None of the participants were carried over from Experiment 1. Participants produced between 2 and 26 (out of 46) Culture Fair errors (M = 11.54, SD = 5.12). Mean Culture Fair error was statistically equivalent to that observed in Experiment 1, $t(46) = -1.68$, $p = .10$; it was therefore unlikely that any differences in the data across experiments were due to differences in *g*.

Materials and Procedure

Materials (the Culture Fair and the colour shape match task) and procedure were carried over from Experiment 1 (see pages 44–45 and 48–49); the only difference was that participants were given four, instead of two, rules to follow at task instructions (although the actual task-related information presented was fundamentally unchanged). The first rule was "respond to items that match in colour by pressing the side corresponding to the placement of the tick." The second rule was "respond to items that match in shape by pressing the side corresponding to the placement of the tick." The third rule was "ignore items that do not match in colour or shape." The fourth rule was "ignore items that match in both colour *and* shape." See Appendix C for the full research protocol.

Design

Performance measures were identical to those in Experiment 1 (see Table 2.1, page 47). A mixed design was adopted in which performance scores and correlations between performance scores and Culture Fair error were compared across performance measures (repeated measures), blocks (repeated measures: inhibition block vs. non-inhibition block), order groups (independent samples: inhibition–non-inhibition group [n = 25] vs. non-inhibition–inhibition group [n = 25]), and rule groups (independent samples: two-rule group [Experiment 1] vs. four-rule group [Experiment 2]).

Results

Main findings

Performance. Table 3.1 compares performance scores across blocks. Critical items were again relatively difficult for participants. Paired t-tests (two-tailed) confirmed that error was significantly greater for critical errors relative to all other measures with the exception of criterion fails in the inhibition block (all $p < .001$, which was significant using the Bonferroni-corrected alpha level of $p < .05 / 9 = .006$). Across

blocks, performance was significantly worse in the inhibition block (relative to the non-inhibition block) for misses, response time, and criterion fails (all $p < .001$, which was significant using the new alpha value of $p < .05 / 6 = .008$; see Table 3.1). These statistical findings were consistent with those observed in Experiment 1.

Table 3.1

Mean performance scores across blocks

	Block						
	Inhibition		Non-inhibition		Paired t-test		
Measure	*M*	*SD*	*M*	*SD*	*t*(49)	*p*	Cohen's *d*
Critical error	.43	.27	-	-	-	-	-
Hand error	.06	.09	.05	.08	.32	.75	.18
Miss	.20	.16	.11	.09	4.68	< .001	.74
False positive	.03	.05	.02	.06	.49	.63	.13
Criterion fail	.49	.27	.14	.13	10.48	< .001	1.54
without CEs	.28	.23	.14	.13	4.33	< .001	.68
Response time	841	93	792	90	5.32	< .001	.75

Note. All means are expressed as proportion of total possible error for that measure with the exception of response time which is presented in ms.

***g* correlations.** Pearson's correlations (one-tailed) between performance scores and Culture Fair (raw error) scores are compared across blocks in Table 3.2. Age was significantly correlated with Culture Fair error, $r(48) = .25$, $p = .04$. As was found in Experiment 1, the performance measures that were significantly correlated with Culture Fair error were not the same as those that were associated with poorer performance. Correlations with Culture Fair error were significant for hand errors and criterion fails in both blocks, for critical errors and misses in the inhibition block only, and with false positives in the non-inhibition block only; significant correlations ranged between $r = .32$ and $r = .65$. Correlations were typically stronger in the inhibition block, and significantly so for the criterion fail measure ($p = .02$). Williams-Hotelling t-tests (two-tailed) also revealed that the correlation with Culture Fair error was significantly higher

for critical errors relative to: false positives in the inhibition block, $t(47) = 2.22$, $p = .03$; response time in the inhibition block, $t(47) = 3.26$, $p = .002$; and response time in the non-inhibition block, $t(47) = 3.15$, $p = .003$. These latter two findings, however, would be expected due to negligible correlations for response time in each block.

Table 3.2

Pearson's correlations between Culture Fair error and performance across blocks

	Block					
	Inhibition		Non-inhibition		William's Hotelling t-test	
Measure	*r*(48)	*p*	*r*(48)	*p*	*t*(47)	*p*
Critical error	.54	< .001	-	-	-	-
Hand error	.40	.002	.33	.01	.79	.43
Miss	.36	.005	.21	.07	1.12	.27
False positive	.17	.12	.32	.01	.74	.46
Criterion fail	.65	< .001	.37	.004	2.32	.02
without CEs	.54	< .001	.37	.004	1.19	.24
Response time	.00	.50	-.02	.45	.19	.85

Culture Fair z-score charts. Participants were assigned to z-score bins (width of .50 SD) based on their Culture Fair error scores. Figures 3.1 and 3.2 display performance scores for each measure across Culture Fair z-score bins. The charts for critical errors, hand errors (both Figure 3.1), and criterion fails (Figure 3.2) demonstrate better performance in participants that fall higher on the *g* distribution. In the inhibition block, participants that scored ≥ 1 SD above the sample mean on the Culture Fair produced 34% fewer critical errors, 21% fewer misses, and 46% fewer criterion fails (but only 3% fewer misses and 8% fewer criterion fails in the non-inhibition block) relative to participants that scored ≥ 1 SD below the sample mean.

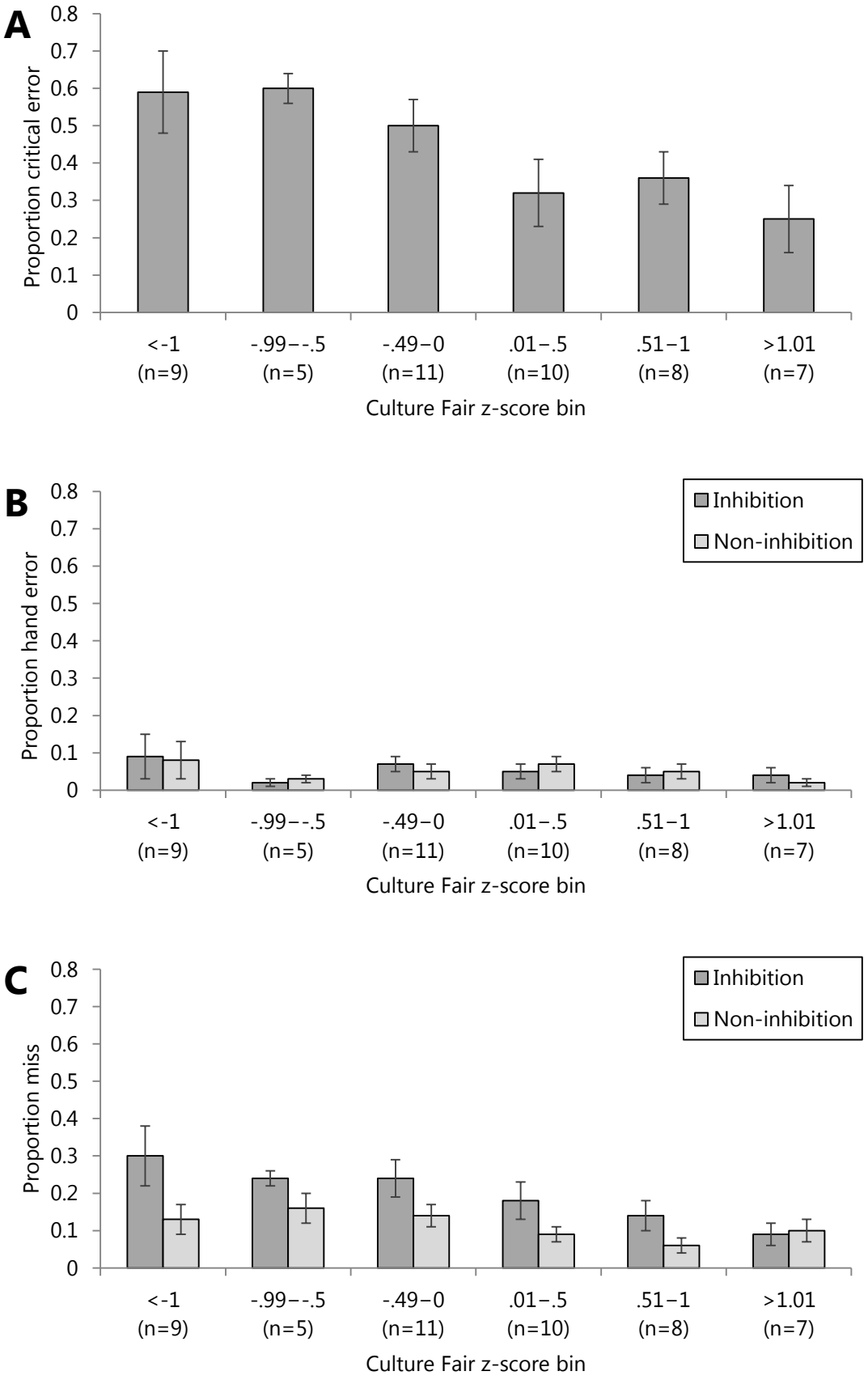


Figure 3.1. Mean performance (proportion error) across Culture Fair z-score bins for critical errors (A), hand errors (B), and misses (C). Error bars represent standard error.

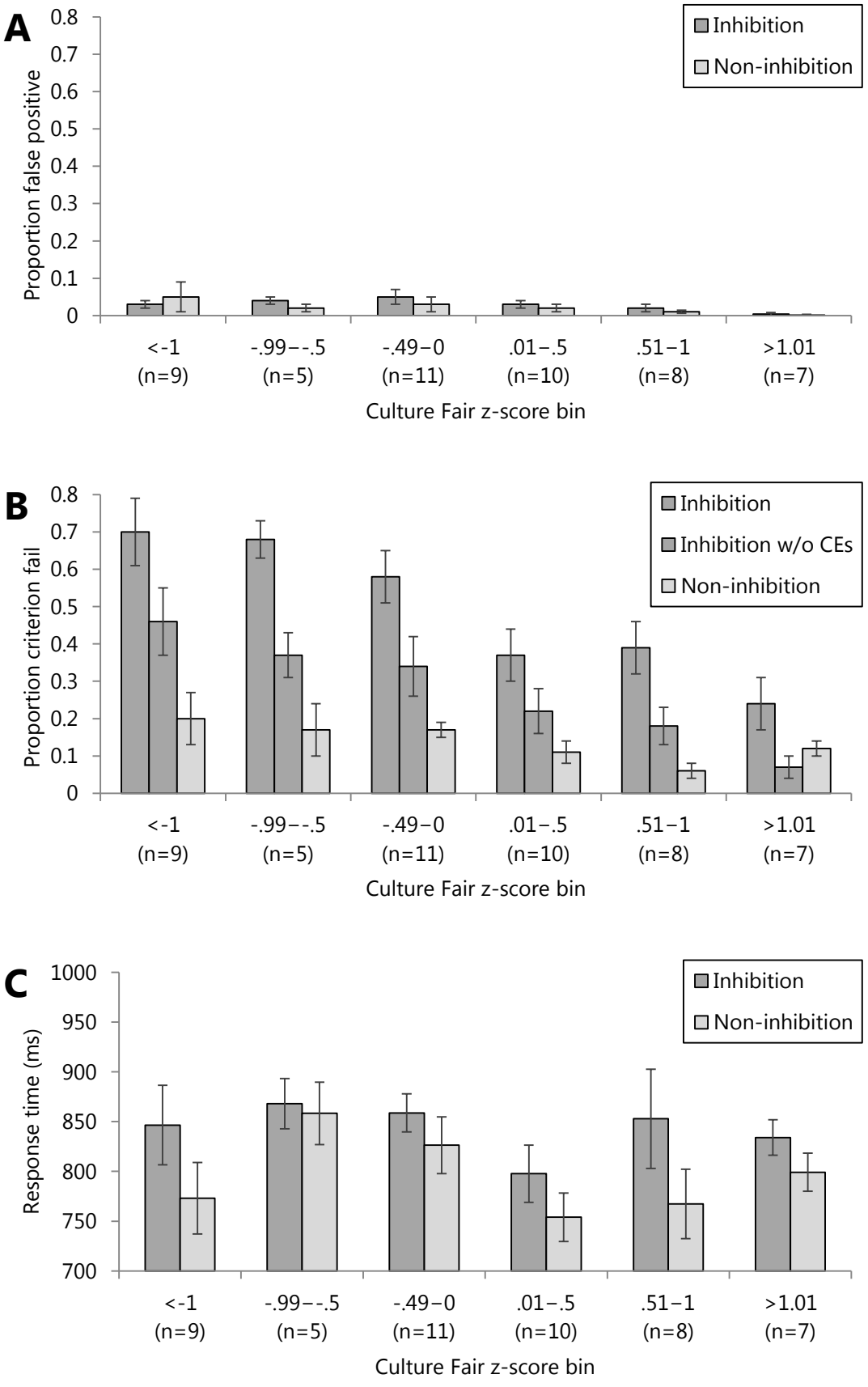


Figure 3.2. Mean performance across Culture Fair z-score bins for false positives (proportion error, A), criterion fails (proportion error, B), and response time (ms, C). Error bars represent standard error.

Comparisons across rule groups. Table 3.3 compares performance levels (independent samples t-tests, two-tailed) and correlations between Culture Fair error and each performance measure (Fisher's z-tests, two-tailed) across rule groups. Fisher's z-test requires a critical value of ±1.96 for significance at the $p < .05$ alpha level. Performance was strikingly similar across the two-rule group of Experiment 1 and the four-rule group of Experiment 2 (see Table 2.2, page 50 and Table 3.1, page 68), and did not differ significantly across groups for any measure. However, correlations between performance and Culture Fair error (see Table 2.3, page 10 and Table 3.2, page 69) were numerically larger in the four-rule group for the majority of measures, and significantly so for criterion fails ($p = .02$); the z-value for critical errors across rule groups was just short of the critical value at 1.95, and the effect for criterion fails without CEs was also marginal ($p = .07$).

Table 3.3

Mean performance scores (independent samples t-tests) and Culture Fair error correlations (Fisher's z-tests) across rule groups

	Independent t-test			Fisher's z-test	
Measure	*t*(98)	*p*	Cohen's *d*	*z*(98)	*p*
		Inhibition block			
Critical error	-.70	.48	-.11	1.95	.05
Hand error	-1.37	.18	-.31	.39	.70
Miss	.08	.94	.00	.59	.56
False positive	1.94	.06	.42	.15	.88
Criterion fail	.09	.93	.04	2.26	.02
without CEs	.51	.61	.08	1.85	.07
Response time	.53	.60	.11	-.29	.77
		Non-inhibition block			
Hand error	-1.77	.08	-.33	.16	.87
Miss	.89	.38	.17	.15	.88
False positive	1.13	.26	.26	.68	.50
Criterion fail	.93	.36	.22	.70	.49
Response time	1.30	.20	.26	.34	.73

Additional findings

Practice effects. ***Performance.*** Performance scores for each sub-block are presented in Table 3.4. There was a general trend for better performance as participants progressed through the task. Table 3.5 shows that repeated measures ANOVAs (two-tailed) revealed a significant improvement in performance across sub-blocks[9] for misses and false positives in both blocks, criterion fails in the inhibition block, and response time in the non-inhibition block (all $p < .01$, except for false positives in the non-inhibition block which was $p = .03$).

Table 3.4

Mean performance scores in each sub-block

	Sub-block 1		Sub-block 2		Sub-block 3	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
			Inhibition block			
Critical error	.50	.33	.43	.36	.37	.37
Hand error	.06	.12	.05	.10	.05	.10
Miss	.27	.21	.16	.19	.17	.17
False positive	.06	.09	.02	.03	.01	.03
Criterion fail	.58	.27	.47	.34	.44	.33
without CEs	.36	.30	.23	.29	.24	.28
Response time	862	131	844	110	824	101
			Non-inhibition block			
Hand error	.07	.10	.05	.08	.04	.09
Miss	.15	.13	.09	.11	.09	.12
False positive	.04	.09	.02	.05	.02	.06
Criterion fail	.17	.17	.13	.19	.12	.19
Response time	818	98	786	101	780	107

[9] Please note that comparisons between sub-blocks have low reliability because each sub-block comprised only four trials.

Table 3.5

Repeated measures ANOVAs comparing mean performance scores across sub-blocks

Measure	*df*	*F*	*p*	η_p^2
		Inhibition block		
Critical error	2, 98	2.60	.08	.05
Hand error	2, 98	.61	.54	.01
Miss [a]	1.63, 79.84	12.73	< .001	.21
False positive [a]	1.23, 60.25	12.5	< .001	.20
Criterion fail	2, 98	6.44	.002	.12
without CEs [a]	1.75, 85.54	5.91	.006	.12
Response time [a]	1.48, 72.66	2.54	.10	.05
		Non-inhibition block		
Hand error	2, 98	1.98	.14	.04
Miss	2, 98	7.13	.001	.13
False positive [a]	1.23, 60.39	4.31	.03	.08
Criterion fail	2, 98	1.53	.22	.03
Response time	2, 98	6.01	.003	.11

Note. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

In line with the findings of Experiment 1, satisfaction of the requirement to inhibit a prepotent response to critical items did not ensure later satisfaction of this requirement. Critical errors across trials for three participants that differed in the amount of Culture Fair error that they produced, alongside the percentage of participants committing critical error on each trial, are presented in Table 3.6. The number of participants making a critical error decreased as the task progressed (from 54% to 32%), but critical error was not prevented by previous successful inhibition of a prepotent response to critical items for 80% of participants. It was again common for participants to notice immediately that they had produced an inappropriate response.

Table 3.6

Pattern of critical error across trials for three participants with differing Culture Fair scores and percentage of participants committing critical error in each trial

Trial	Pattern of critical error			% of participants
	Participant A (3 Culture Fair errors)	Participant B (8 Culture Fair errors)	Participant C (18 Culture Fair errors)	
1	0	0	1	54
2	1	1	1	50
3	0	0	1	44
5	0	0	1	48
7	0	1	0	46
8	0	1	0	32
10	0	1	1	40
11	0	0	1	40
12	0	1	0	32
Overall	1	5	6	80

g correlations. Pearson's correlations (one-tailed) between sub-block performance scores and Culture Fair error are presented in Table 3.7. In the inhibition block, Culture Fair error was significantly correlated with critical errors, misses and criterion fails in every sub-block, and with hand errors in sub-blocks 1 and 2. In the non-inhibition block, Culture Fair error was significantly correlated with false positives and criterion fails in every sub-block, with misses in sub-block 1, and with hand errors in sub-blocks 2 and 3. Williams-Hotelling t-test (two-tailed) revealed that the correlation with Culture Fair error for hand errors was significantly stronger: in sub-block 2 relative to sub-block 1 in the inhibition block, $t(47) = 2.46$, $p = .02$; in sub-block 2 relative to sub-block 3 in the inhibition block, $t(47) = 2.20$, $p = .03$; and in sub-block 3 relative to sub-block 1 in non-inhibition block, $t(47) = 2.37$, $p = .02$. For the critical error correlation, the Williams-Hotelling t-test between sub-blocks 1 and 3 reached 1.98 (with a numerically higher correlation in sub-block 3), which is just short of the critical value of ±2.01 required for a significant result.

Table 3.7

Pearson's correlations between Culture Fair error and sub-block performance

	Sub-block 1		Sub-block 2		Sub-block 3	
Measure	*r*(48)	*p*	*r*(48)	*p*	*r*(48)	*p*
			Inhibition block			
Critical error	.26	.04	.40	.002	.56	< .001
Hand error	.26	.04	.52	< .001	.23	.06
Miss	.38	.003	.24	.05	.27	.03
False positive	.20	.08	.06	.33	.10	.24
Criterion fail	.57	< .001	.49	< .001	.59	< .001
without CEs	.53	< .001	.36	.005	.39	.002
Response time	.06	.34	-.07	.31	.004	.49
			Non-inhibition block			
Hand error	.14	.18	.26	.03	.45	.001
Miss	.24	.05	.12	.21	.13	.19
False positive	.25	.04	.31	.02	.32	.01
Criterion fail	.26	.03	.29	.02	.24	.04
Response time	-.03	.43	.02	.46	-.06	.34

Comparisons across rule groups. For each sub-block measure, performance scores and correlations with Culture Fair error were compared across rule groups. Independent samples t-tests (two-tailed) showed that error was significantly greater in the two-rule group relative to the four-rule group for criterion fails in sub-block 1 of the non-inhibition block[10], $t(74.77) = 2.05$, $p = .04$, $d = -.43$; the same pattern of error was observed for false positives in sub-block 2 of the inhibition block[10], $t(51.20) = 2.17$, $p = .04$, $d = -.25$. All other performance comparisons were non-significant (all at $p > .10$), but the effects for misses, false positives [each in sub-block 1 of the non-inhibition block; each more error-prone in two-rule], and hand errors [in sub-block 3 of the inhibition block; more error prone in four-rule] were just short of significance ($p = .053$ to .054). Fisher's z-tests (two-tailed) revealed that correlations with Culture Fair error were significantly stronger in the four-rule group relative to the two-rule group for: critical errors in sub-block 2, $z(98) = 2.25$, $p = .03$; criterion fails in sub-block 3 of the

[10] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

inhibition block, $z(98) = 2.00$, $p < .05$; and hand errors in sub-block 3 of the non-inhibition block, $z(98) = 1.96$, $p < .05$. All other comparisons were non-significant ($p > .10$).

Order effects. ***Performance.*** Table 3.8 compares performance scores across order groups. As was found in Experiment 1, there was a trend for greater error in the block that was completed first. However, a greater number of statistical differences across order groups were observed here. Independent samples t-tests (two-tailed) showed that error was greater in the inhibition block for the inhibition–non-inhibition group (relative to the non-inhibition–inhibition group) for critical errors ($p = .02$), misses ($p = .001$), false positives ($p = .006$), and criterion fails ($p = .001$ to $.008$). This was despite the observation of statistically equivalent Culture Fair scores across order groups.

Table 3.8

Mean performance scores across order groups

	Order group							
	Inhibition–non-inhibition		Non-inhibition–inhibition		Independent t-test			Cohen's
Measure	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	*d*
Culture Fair	12.12	4.56	10.96	5.66	48	.80	.43	.23
			Inhibition block					
Critical error	.52	.22	.35	.29	48	2.40	.02	.67
Hand error	.06	.06	.05	.11	48	.73	.47	.12
Miss	.27	.16	.13	.13	48	3.44	.001	.97
False positive [a]	.05	.06	.01	.02	27.30	3.00	.006	1.00
Criterion fail [a]	.61	.19	.38	.28	42.26	3.41	.001	.98
without CEs	.36	.22	.19	.21	48	2.77	.008	.79
Response time	865	72	817	107	48	1.87	.07	.54
			Non-inhibition block					
Hand error	.04	.05	.07	.09	48	-1.17	.25	-.43
Miss	.11	.11	.11	.07	48	.08	.94	.00
False positive [a]	.01	.01	.04	.09	24.78	-1.71	.1	-.60
Criterion fail	.14	.13	.14	.12	48	.03	.97	.00
Response time	806	85	778	94	48	1.11	.27	.31

Note. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

g correlations. Table 3.9 compares Pearson's correlations (one-tailed) between performance and Culture Fair error across order groups. Correlations with Culture Fair error appeared larger in the non-inhibition–inhibition group for the majority of measures. Fisher's z-test (two-tailed) revealed that these differences were significant for critical errors ($p = .03$) and hand errors in both blocks ($p = .003$ to.005). The correlation between response time (in the non-inhibition block) and Culture Fair error was significantly stronger in the inhibition–non-inhibition group ($p = .04$), but this correlation was actually negative in the non-inhibition–inhibition group.

Table 3.9

Pearson's correlations between Culture Fair error and performance across order groups

	Order group					
	Inhibition–non-inhibition		Non-inhibition–inhibition		Fisher's z-test	
Measure	*r*(23)	*p*	*r*(23)	*p*	*z*(48)	*p*
			Inhibition block			
Critical error	.27	.10	.74	< .001	2.23	.03
Hand error	-.06	.38	.69	< .001	3.01	.003
Miss	.41	.02	.20	.20	.77	.44
False positive	.16	.22	.11	.32	.17	.87
Criterion fail	.66	< .001	.68	.001	.12	.90
without CE	.57	.002	.55	.006	.10	.92
Response time	.16	.22	-.30	.10	-1.56	.12
			Non-inhibition block			
Hand error	-.01	.48	.68	< .001	2.78	.005
Miss	.25	.12	.15	.27	.35	.73
False positive	.14	.26	.62	.002	.19	.06
Criterion fail	.19	.18	.49	.02	1.14	.25
Response time	.25	.11	-.35	.07	-2.06	.04

Note. $n = 25$ in each order group.

Comparisons across rule groups. Performance scores and Culture Fair correlations in each order group were compared with those observed in the order groups of Experiments 1. Independent t-tests (two-tailed) showed that performance did not differ significantly across rule groups for any of the order groups ($p > .09$ for all measures). However, Fisher's z-tests revealed that, in non-inhibition–inhibition participants, the correlation with Culture Fair error was significantly stronger in the four-rule group relative to the two-rule group for critical errors, $z(98) = 2.16$, $p = .03$; the same pattern was observed for criterion fails in the inhibition block, $z(98) = 2.10$, $p = .04$. All other comparisons were non-significant (all $p > .08$).

Learning and reconceptualisation of task rules. Using their description of task rules (which was recorded at three points during the experiment: at task instructions, after the first block, and after the second block), participants were categorised as either (a) being correct (if they stated all rules without any omissions, not necessarily word-for-word, in the first instance) or incorrect (if they omitted parts of rules in the first instance), and (b) having reconceptualised (if they stated fewer or more than four distinct rules) or not reconceptualised (if they stated four distinct rules). An independent-samples t-test (two-tailed) revealed that Culture Fair error was significantly lower in participants that stated the rules correctly (n = 36, M = 10.47 errors, SD = 4.59) relative to those that stated the rules incorrectly (n = 14, M = 14.29 errors, SD = 5.55) after the first block, $t(48) = 2.49$, $p = .02$, $d = .43$; all other comparisons were non-significant ($p < .10$).

Rule failure scores and total proportion error scores. Participants were assigned rule failure and proportion error scores for each block, based on the methods for data analysis in Duncan et al. (2008); see page 60 for a description of how these rule failure and proportion error scores were calculated. Table 3.10 compares rule failure scores, total proportion error scores (using analysis of covariance [ANCOVAs], controlling for Culture Fair error[11]), and their respective correlations with Culture Fair error (using Fisher's z-tests) across rule groups. For the four-rule data, paired t-tests (two-tailed) revealed that rule failure scores were significantly higher in the inhibition block relative to the non-inhibition block[12], $t(49) = 4.70$, $p < .001$; the same effect was observed for proportion error scores, $t(49) = 8.55$, $p < .001$, $d = 1.32$. Williams-Hotelling t-test (two-tailed) showed that the correlation with Culture Fair error for the proportion error score was statistically equivalent across blocks ($p > .10$). The correlations with Culture Fair

[11] Note that ANCOVA was used for comparisons across rule groups for rule failure and proportion error scores in order to be consistent with the analyses conducted in Duncan et al (2008); Culture Fair error was not statistically controlled for in other comparisons across rule groups because the groups did not differ in Culture Fair error.

[12] Cohen's *d* was not computed because the mean rule failure score in the non-inhibition block was zero.

error for the rule failure scores were not compared across blocks due to the observation of ceiling effects in the non-inhibition block.

Across rule groups, ANCOVAs (one-tailed[13]) revealed that both rule failure and total proportion error scores were statistically equivalent across the two-rule group and the four-rule group. However, Fisher's z-test (one-tailed) revealed that the correlations with Culture Fair error for both rule failure and proportion error scores were significantly stronger in the four-rule group relative to the two-rule group for the inhibition block and for both blocks combined.

Table 3.10

Rule failure and total proportion error scores across rule groups

	Rule group				ANCOVA		
	Two-rule		Four-rule				
Block	*M*	*SE*	*M*	*SE*	$F(1, 97)$	*p*	η_p^2
			Rule failure score				
Inhibition	.49	.11	.55	.11	.13	.72	.001
Non-inhibition	.15	.06	.01	.06	3.39	.07	.03
Both blocks	.64	.13	.54	.13	.29	.59	.003
			Total proportion error score				
Inhibition	.25	.02	.23	.02	.90	.34	.01
Non-inhibition	.12	.02	.09	.02	1.41	.24	.01
Both blocks	.18	.01	.16	.01	1.67	.20	.02

[13] One-tailed tests were employed here due to the observed differences between task model complexity groups in Duncan et al. (2008)

Table 3.11

Pearson's correlations with Culture Fair error for rule failure and total proportion error scores across rule groups

	Rule group					
	Two-rule		Four-rule		Fisher's z-test	
Block	*r*(48)	*p*	*r*(48)	*p*	*z*(98)	*p*
			Rule failure score			
Inhibition	.21	.07	.51	< .001	1.65	< .05
Non-inhibition	.13	.18	-	-	-	-
Both blocks	.22	.07	.51	< .001	1.64	.05
			Total proportion error score			
Inhibition	.34	.01	.62	< .001	2.58	.01
Non-inhibition	.28	.02	.46	< .001	1.46	.07
Both blocks	.35	.006	.66	< .001	3.06	.001

Note. A comparison of rule failure scores in the non-inhibition block across rule groups was not possible because four-rule participants did not produce any rule failures in this block.

Discussion

Despite statistically equivalent performance across experiments, the correlation with *g* was significantly greater in the four-rule group (of Experiment 2) relative to the two-rule group (of Experiment 1) for the two central measures (critical errors and criterion fails; i.e., the measures that are associated with response inhibition). Like Experiment 1, greater error was associated with the presence of a requirement to inhibit a prepotent but inappropriate response tendency: (a) greater error was produced on critical items relative to other measures; (b) successful inhibition of a response to critical items did not prevent later critical error for 80% of the sample; and (c) performance was significantly worse in the inhibition block relative to the non-inhibition block for misses, response time, and criterion fails. In as much as four-rule instructions represented a more complex set of instructions, the statistically different correlations between task performance and Culture Fair error across experiments supports prior research. Duncan et al. (2008) also observed increased correlations between *g* and goal

neglect when a full, rather than a reduced, set of instructions was presented to participants; unlike here, however, they also found significantly worse performance in the full-instructions group.

In Experiment 2 only, the correlation with *g* for performance on items requiring the inhibition of a prepotent but inappropriate response tendency reached significance, and it was also significantly stronger than the *g* correlations for other measures that did not require response inhibition (specifically, false positives in the inhibition block and response time in both blocks). This finding of a significant relationship between response inhibition and *g* is consistent with previous research that has also found such a relationship (e.g., Polderman et al., 2009). The observation of a significant response inhibition–*g* correlation in the context of four task rules, but not in the context of two task rules, may explain the conflicting evidence for an inhibition–*g* relationship in the literature. Such disagreement may be a result of failing to control for the complexity of task instructions and thus failing to control for the load imposed on the working memory system by task instructions. This view, however, is quite difficult to validate with precision because very few published studies provide verbatim participant instructions.

Differences in performance–*g* correlations between order groups, which were not observed in Experiment 1, were also observed here. The correlation with *g* for critical errors was significantly stronger in the non-inhibition–inhibition order group relative to the inhibition–non-inhibition order group (as were the *g* correlations for hand errors in both blocks and response time in the non-inhibition block). The requirement to ignore double-matching frames was presented as a part of a rule (rule 2) in two-rule instructions, but was presented in isolation (rule 4) in four-rule instructions. If the ability to use task rules is related to *g*, perhaps these order group findings can be explained by the greater ability of people at the higher end of the *g* distribution to adhere to a rule that has not previously been required for correct performance; by completing the 12 trials of the non-inhibition block (i.e., a task that did not contain

double-matching items) prior to attempting the inhibition block, the response inhibition requirement had arguably not been enforced for non-inhibition–inhibition participants. Conversely, inhibition–non-inhibition participants used, and therefore enforced, the double-match requirement straight away.

Indeed, another finding that is suggestive of a relationship between response inhibition and *g* is that the overall error–*g* correlation was significantly stronger in the inhibition block relative to the non-inhibition block. However, because error was not statistically equivalent across blocks it is unclear whether this finding is truly indicative of a relationship between response inhibition and *g*, or whether it can instead be explained on the basis of task difficulty. Another interpretation of this result could be that the greater demand for operating/executing a task rule that had not been enforced in four-rule participants may explain why the correlation with overall error was significantly stronger in the inhibition block (compared to non-inhibition block) in four-rule participants only.

At first glance, these findings suggest that when memory load for task rules is greater, the recruitment of *g* in task performance increases. However, not only is memory load arguably higher when instructions are given as four (compared to two) rules, but the actual learning of task requirements, or the *formation* of the mental representation of the task may have been more difficult. Key to success on the task was to understand the importance of matching on a single dimension. This understanding may happen early in two-rule instructions because the first rule requires participants to respond to items matching in "colour *or* shape," but may happen much later in four-rule instructions (at the presentation of rule 4) due to the removal of this operative word *or* from the go rules (rules 1 and 2). The need to update or reconceptualise the task conceptualisation in working memory may therefore be present in four-rule instructions, but absent from two-rule instructions, because the understanding of the task formed on the basis of previously specified rules is altered in four-rule instructions only.

In conclusion, the findings of Experiments 1 and 2 suggest that the way in which a task is cognitively modelled is a critical factor in the recruitment of *g* in task performance (in conditions of response inhibition). Perhaps when a larger amount of working memory "space" is used by the representation and use of task instructions (in the four-rule, relative to the two-rule, condition), additional recruitment of *g* is required in order to manage the performance requirements of the task. When such working memory space is left unfilled (in two-rule instructions) the system may have adequate resources to cope with the processing demands of the task without recruiting *g*. In Experiment 3, instructions for the same task were again presented as four rules; however, the requirement to reconceptualise the task was removed from the process of learning task rules in order to determine whether this, rather than the number of rules, was driving the patterns in the data.

CHAPTER 4

Reconceptualisation of Task Requirements

Overview

Chapter 4 explored two issues with respect to task model *reconceptualisation*: (a) whether the increase in task-sensitivity to *g*, observed in Experiment 2, was due to the addition of a requirement to alter the task conceptualisation, formed in the context of prior rules, as later rules were presented, rather than an increase in the number of distinct rules presented; and (b) whether the way in which participants represent the task in mind varies as a function of *g*. The method of Experiment 3 was carried over from Experiment 2 but task instructions were modified such that the importance of matching on a single dimension was flagged early (rather than late). Performance scores were statistically equivalent across Experiments 1 through 3. Critical error and overall error were significantly correlated with *g*, consistent with observations in the four-rule group (Experiment 2). Unlike Experiment 2, these correlations were not significantly stronger to those found in the two-rule group (Experiment 1), and the correlation between overall error and *g* was statistically equivalent across the inhibition and non-inhibition blocks. A second modification to the method involved prompting participants, post-execution, to verbalise the rules "as they are represented in your own mind." Those participants that reconceptualised the rules tended to score well on the Culture Fair. These findings suggest that the more crucial aspect of task model complexity in increasing task-sensitivity to *g* is the number of rules, but that the ability to reconceptualise task rules is nonetheless linked to *g*.

Experiment 3

The results from Experiment 2 showed that task-sensitivity to Spearman's *g* increases as the number of rules presented at task instructions increases, particularly

when the task demands the inhibition of a prepotent response tendency. This was despite controlling for the amount of operative task information presented, but by manipulating instead how this information was chunked into distinct rules. This finding appears to be independent of task difficulty as demonstrated by statistically equivalent performance scores (error and response time) across two-rule and four-rule groups. Therefore, some factor(s) other than increased task difficulty must be strengthening the correlations between task performance and *g*. Instead of assuming that the increased *number* of presented task rules in the four-rule condition is responsible for the stronger *g* correlations, we can address the issue of differences in the process of *forming* the task model between two-rule and four-rule conditions.

A consequence of splitting the two rules of Experiment 1 into the four rules of Experiment 2 may have been the addition of a requirement to alter the understanding of the task during the process of forming the task model. It can be argued that, in the presentation of two-rule instructions, the importance of matching on a single dimension is indicated early-on by the inclusion of the logical operator *or* in the first rule ("respond to items that match in colour *or* shape by…"). However, in the presentation of four-rule instructions, the process of separating out this information into two parts removes this operator ("respond to items that match in colour by…"). This has the effect of implying that a response should be made to items that match, irrespective of the number of matching dimensions. Consequently, the understanding of the task is altered later-on in four-rule instructions at the presentation of the final rule, which indicates the importance of responding to matches on a single dimension only ("ignore items that match in colour *and* shape"). Perhaps some of kind of cognitive flexibility is required to incorporate this new understanding into the conceptualisation of the task, a capacity that is not necessary in the formation of the two-rule task conceptualisation.

The ability to alter the understanding of the task during the formation of the task model could tap into an inhibitory function that may indeed be associated with *g*. It is possible that participants that fall lower on the *g* distribution find it more difficult to

clarify/alter/update their understanding of previously specified rules on the basis of a new task constraint. People with low *g*, relative to people with high *g*, could be more "stuck in" to a somewhat inflexible task model, finding it more difficult to ignore the previous model of the task and attend to the altered model of the task. This is suggested by the perseveration behaviours displayed by people with frontal lesions (e.g., Luria, 1966), coupled with the association between frontal lobe function and fluid intelligence (e.g., Barbey, Colom, & Grafman, 2012). On these grounds, any reduction of task-sensitivity to *g* due to the removal of the requirement to alter the rule boundaries (from four-rule instructions) may support the importance of inhibitory functions to intelligence.

Another aspect of task model reconceptualisation was highlighted by the methods of Experiments 1 and 2: when repeating task instructions subsequent to task execution, some participants split the information into fewer or more chunks than was set out in the task instructions that were initially presented to them. A comparison of Culture Fair scores between participants that did, and participants that did not, reconceptualise task rules revealed no statistical differences. However, participants received no explicit instruction to impose their own order on task instructions, so it is possible that those participants that may have reconceptualised the rules were not identified. Thus, in Experiment 3, participants were asked to state task rules as they are were mentally represented. Such an investigation more clearly defines the relationship, if such a relationship exists, between reconceptualisation of task requirements and *g*. Perhaps people that reconceptualise task rules in an efficient way (i.e., reduce the number of chunks represented within the task model) perform better on the task due to using a more ordered model to effectively control behaviour.

The purpose of Experiment 3 was thus two-fold. One aim was to address whether the number of specified rules, or an added requirement to alter the format of the task model whilst learning task rules, was responsible for increasing the strength of correlations between task performance and *g*. The other aim was to investigate whether

it is predominantly higher-*g* participants that reconceptualise task rules forming a more efficient task model than was encouraged by their received instructions. Accordingly, the method of Experiment 2 was modified in two ways in Experiment 3. The first modification was the addition of the word *only* to the first two rules of four-rule instructions (e.g., "respond to items that match *only* in colour by…"). It was therefore assumed that the importance of matching on a single dimension would be flagged early-on in task instructions, eliminating the need to reconceptualise the task during the learning of task rules. The second modification was to explicitly instruct participants to verbalise the rules "in the way that best reflects how the rules are represented in your *own mind*" in order to more reliably detect participants that may form flexible task representations. Performance scores and correlations between performance and scores on the Culture Fair were compared with those observed in previous experiments. Thus, how the relationship between inhibition and *g* is affected by the number of rules presented at task instructions was explored whilst controlling for a requirement to form a flexible task model.

It was unclear what effect, if any, the addition of the word *only* to four-rule instructions, would have on the recruitment of *g*. Such a simple manipulation to the format of the administered instructions may be too minute to elicit any observable differences. Nonetheless, on the assumption that a body of four rules increases the involvement of *g* (relative to a body of two rules) it was predicted that correlations with Culture Fair error would be stronger in the four-rule-ONLY group (Experiment 3) relative to the two-rule group (Experiment 1). If task reconceptualisation adds to the recruitment of *g* then any correlations observed here may not be as strong as those observed in Experiment 2. On these grounds, it was also predicted that participants that reconceptualised the task efficiently (i.e., stated fewer than four rules when asked to state the rules as they were represented in mind) would produce fewer Culture Fair errors relative to participants that did not reconceptualise task rules.

Method

Participants

An opportunity sample of 20 adults (15 female), aged between 18 and 53 years (M = 24.00 years, SD = 10.00), and with no history of neurological disorder, was recruited for Experiment 3. None of the participants were carried over from previous experiments. Participants consisted of undergraduate students from the Department of Psychology, Anglia Ruskin University and community volunteers. Participants produced between 2 and 24 Culture Fair errors (M = 9.10, SD = 5.82). Mean Culture Fair error was statistically equivalent to that observed in Experiment 1, $t(68) = .72$, $p = .48$, $d = -.17$; Culture Fair error was also equivalent to that observed in Experiment 2, $t(68) = 1.73$, $p = .09$, $d = -.45$.

Materials and Procedure

Materials and procedure were carried over from Experiment 2, with the exception of two modifications. The first modification was the inclusion of the word *only* in Rules 1 and 2. The first rule was "respond to items that match *only* in colour by pressing the side corresponding to the placement of the tick." The second rule was "respond to items that match *only* in shape by pressing the side corresponding to the placement of the tick." The third rule was "ignore items that do not match in colour or shape" (unchanged from Experiment 2). The fourth rule was "ignore items that match in both colour *and* shape" (unchanged from Experiment 2). The second modification involved participants stating the rules as they were represented in their own mind, rather than simply repeating the rules after task execution. Specifically, participants were asked to imagine that they had swapped roles with the experimenter and that they were administering the task and its instruction rules to someone else. Participants were reminded of the four rules that were presented at task instructions, and were then asked to think about whether this information was represented in their mind in the same, or a different, way to the initial explanation of the rules. In their own time, participants

verbalised the content of, and how many rules were in, their own task conceptualisation. See Appendix D for the full research protocol.

Design

Performance measures were identical to those in Experiments 1 and 2 (see Table 2.1, page 47). A mixed design was adopted in which performance scores and correlations between performance scores and Culture Fair error were compared across performance measures (repeated measures), blocks (repeated measures: inhibition vs. non-inhibition), order groups (independent samples: inhibition–non-inhibition group [n = 10] vs. non-inhibition–inhibition group [n = 10]), and rule groups (independent samples: four-rule-ONLY group [Experiment 3] vs. two-rule group [Experiment 1] and four-rule group [Experiment 2]).

Results

Main findings

Performance. Performance scores are compared across blocks in Table 4.1. Consistent with previous experiments, critical items were relatively difficult for participants. Paired t-tests (two-tailed) showed that error was significantly greater for critical errors than for the majority of measures (i.e., all except misses and criterion fails in the inhibition block; all $p < .004$, which was significant using the Bonferroni-corrected alpha level of $p < .05 / 9 = .006$). Across blocks, performance was significantly worse in the inhibition block relative to the non-inhibition block for criterion fails (when using critical error as a criterion; $p < .001$, significant using the new alpha value of $p < .05 / 6 = .008$).

Table 4.1

Mean performance scores across blocks

	Block						
	Inhibition		Non-inhibition		Paired t-test		
Measure	*M*	*SD*	*M*	*SD*	*t*(19)	*p*	Cohen's *d*
Critical error	.39	.30	-	-	-	-	-
Hand error	.04	.03	.04	.05	-.17	.87	.00
Miss	.17	.13	.14	.13	1.69	.11	.24
False positive	.03	.04	.02	.03	.98	.34	.40
Criterion fail	.42	.22	.15	.16	4.25	< .001	.99
without CEs	.21	.16	.15	.16	2.02	.06	.44
Response time	849	93	857	112	-.39	.70	-.09

Note. All means are expressed as proportion of total possible error for that measure except for reaction time which is presented in ms.

***g* correlations.** Pearson's correlations (one-tailed) between performance scores and Culture Fair error are presented in Table 4.2. For a sample size of 20, the critical value for Pearson's *r* (one-tailed) is ±.38 at the $p < .05$ alpha level. Correlations with Culture Fair error were significant for critical errors, misses, false positives, and criterion fails in the inhibition block, and for response time in both blocks; significant correlations ranged between $r = .40$ and $r = .76$. Williams-Hotelling t-tests (two-tailed), which require a critical value of ±2.10 for significance at the $p < .05$ alpha level for a sample size of 20, showed that the correlation with Culture Fair error for critical errors did not differ in strength from that for any other performance measure (all $p > .10$). However, correlations were significantly higher in the inhibition block relative to the non-inhibition block for misses ($p < .05$), false positives ($p = .03$), and criterion fails (when determined without critical error, $p < .001$). The correlation between age and Culture Fair error was marginal, $r(18) = .37$, $p = .053$.

Table 4.2

Pearson's correlations between Culture Fair error and performance across blocks

	Block					
	Inhibition		Non-inhibition		William's Hotelling t-test	
Measure	*r*(18)	*p*	*r*(18)	*p*	*t*(47)	*p*
Critical error	.49	.01	-	-	-	-
Hand error	.11	.32	.20	.20	0.39	.70
Miss	.64	.001	.25	.15	2.13	< .05
False positive	.64	.001	.32	.08	2.48	.03
Criterion fail	.66	.001	.25	.14	1.41	.18
without CEs	.76	< .001	.25	.14	4.21	< .001
Response time	.58	.004	.40	.04	1.02	.32

Culture Fair z-score charts. Figures 4.1 and 4.2 display performance scores for each measure across Culture Fair z-score bins (width of .50 SD). The charts for critical errors, misses (both Figure 4.1), false positives, and criterion fails (both Figure 4.2) demonstrate reduced error as a function of Culture Fair performance, and the chart for response time (Figure 4.2) demonstrates faster speed as a function of Culture Fair performance. Participants that scored ≥ 1 SD above the sample mean on the Culture Fair produced 41% fewer critical errors, 17% fewer misses, and 42% fewer criterion fails in the inhibition block (but only 12% fewer misses and 4% fewer criterion fails in the non-inhibition block) relative to participants that scored ≥ 1 SD below the sample mean.

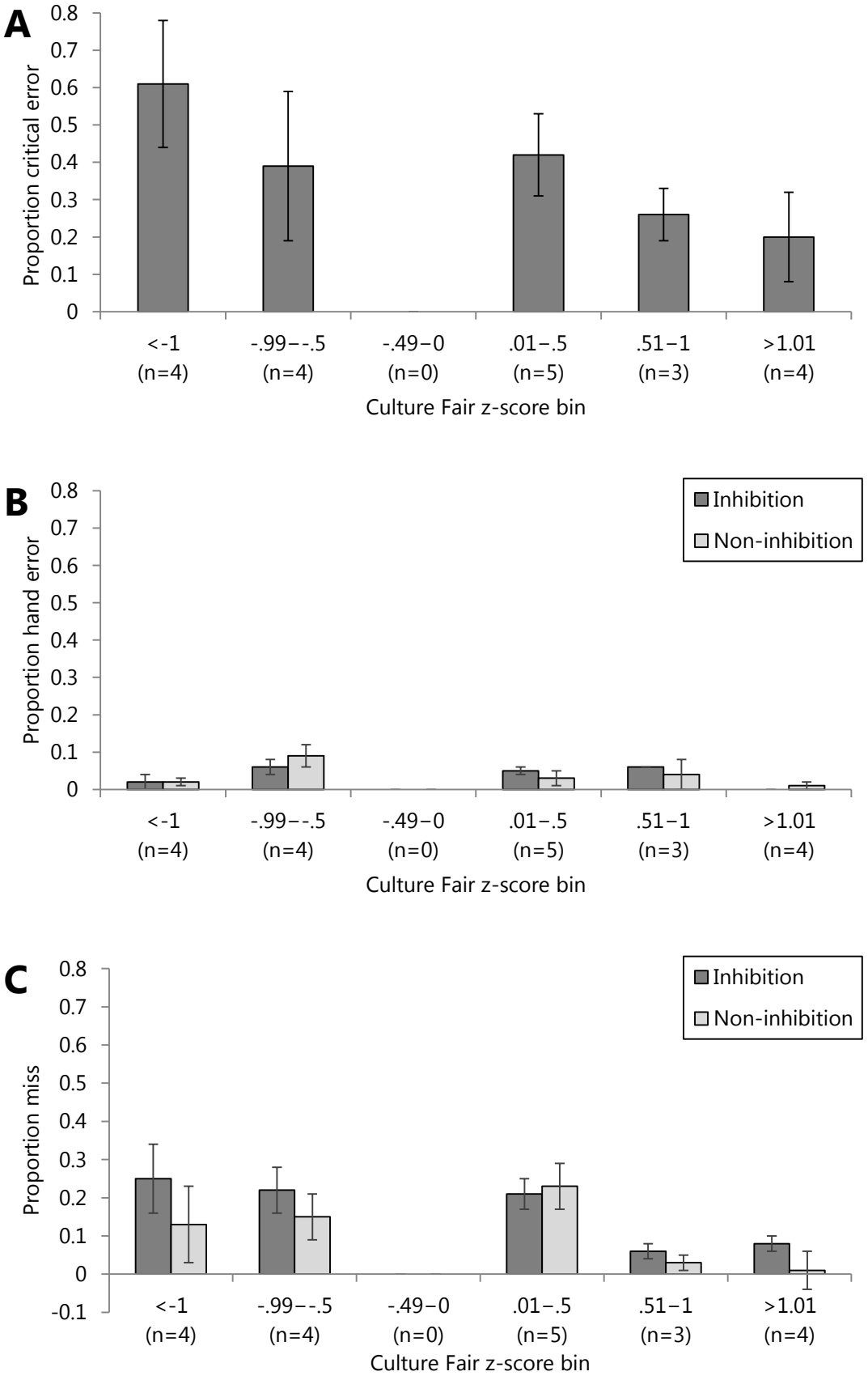


Figure 4.1. Mean performance (proportion error) across Culture Fair z-score bins for critical errors (A), hand errors (B), and misses (C). Error bars represent standard error.

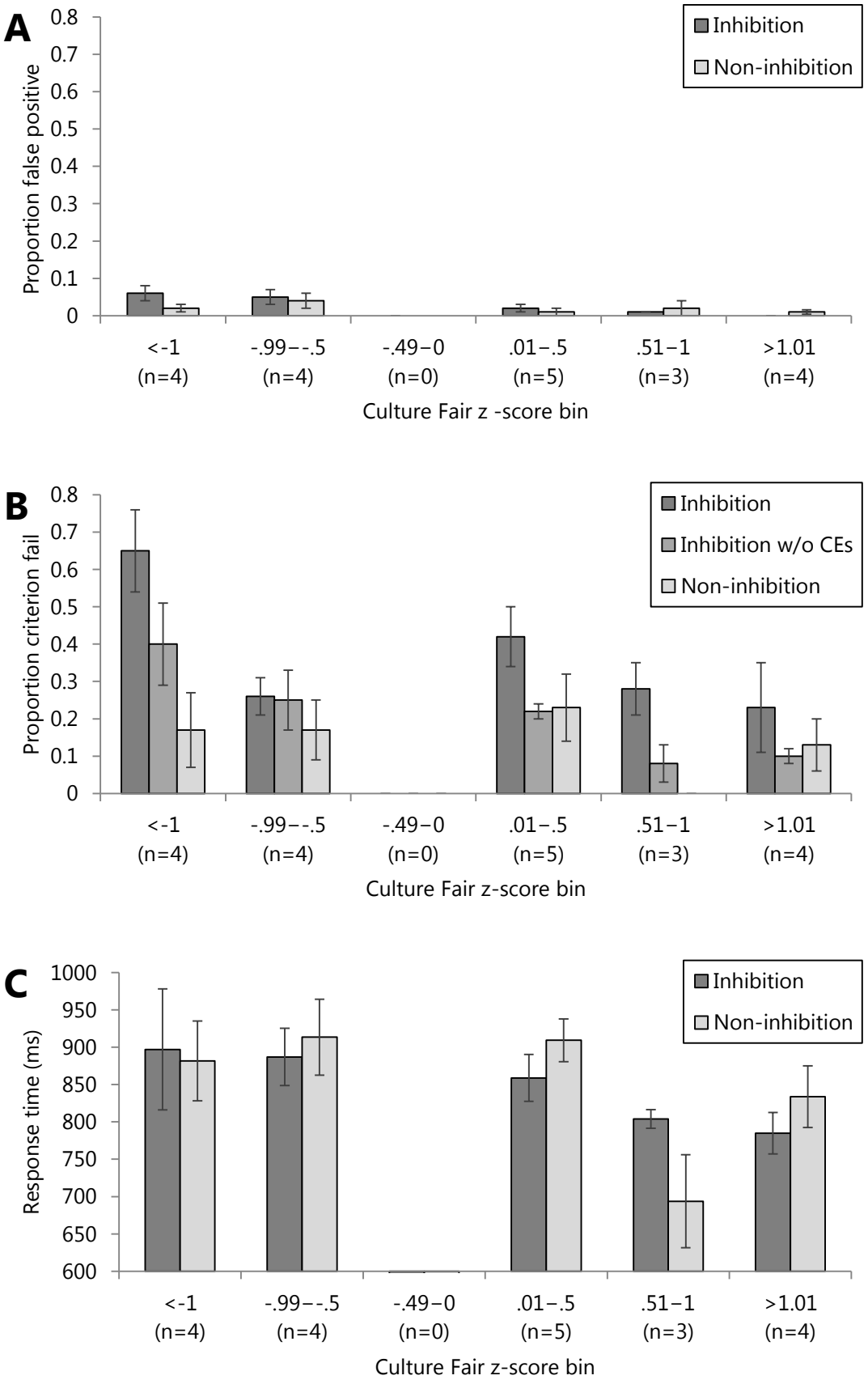


Figure 4.2. Mean performance across Culture Fair z-score bins for false positives (proportion error, A), criterion fails (proportion error, B), and response time (ms, C). Error bars represent standard error.

Comparisons across rule groups. Table 4.3 compares performance scores (independent samples t-tests, two-tailed) and correlations between Culture Fair error and each performance measure (Fisher's z-tests, two-tailed) across rule groups. The only performance measure that differed across rule groups was response time, which was significantly slower in Experiment 3 (four-rule-ONLY) relative to Experiment 2 (four-rule) in the non-inhibition block ($p = .03$). A number of correlations between performance and Culture Fair error were significantly higher in Experiment 3 relative to previous experiments. The correlation with Culture Fair error was significantly higher in Experiment 3 relative to Experiment 1 for false positives ($p = .03$), criterion fails (determined without critical error, $p = .01$), and response time ($p = .01$); correlations with Culture Fair error for false positives ($p = .04$) and response time ($p = .02$; all in the inhibition block only) were also significantly stronger in Experiment 3 relative to Experiment 2.

Table 4.3

Mean performance scores (independent samples t-tests) and Culture Fair correlations (Fisher's z-tests) across rule groups

Measure	Independent t-test		Cohen's *d*	Fisher's z-test	
	t(68)	*p*		*z*(68)	*p*
(a) Experiment 3 (four-rule-ONLY) vs. Experiment 1 (two-rule)					
Inhibition block					
Critical error	.09	.93	.04	1.18	.24
Hand error	.16	.87	.00	.82	.41
Miss	.82	.42	.21	1.85	.06
False positive	1.32	.19	.44	2.18	.03
Criterion fail	1.20	.23	.33	1.75	.08
without CEs	1.32	.19	.41	2.73	.01
Response time	.05	.96	.01	2.55	.01
Non-inhibition block					
Hand error	-.49	.62	-.22	.38	.71
Miss	-.12	.91	-.07	.00	1.00
False positive	.89	.38	.30	.49	.62
Criterion fail	.43	.67	.15	.04	.97
Response time	-1.55	.13	-.40	1.32	.19
(b) Experiment 3 vs. Experiment 2 (four-rule)					
Inhibition block					
Critical error	.60	.55	.14	.24	.81
Hand error	1.01	.32	.33	1.11	.27
Miss	.76	.45	.21	1.35	.18
False positive	.27	.79	.00	2.04	.04
Criterion fail	1.13	.26	.29	.06	.95
without CEs	1.09	.28	.36	1.39	.17
Response time	-.34	.74	-.09	2.34	.02
Non-inhibition block					
Hand error	.88	.38	.15	.50	.62
Miss	-.90	.37	-.27	.15	.88
False positive	.35	.73	.00	.04	.97
Criterion fail	-.28	.78	-.07	.47	.64
Response time	-2.32	.03	-.64	1.56	.12

Additional findings

Practice effects. ***Performance.*** Performance scores for each sub-block are presented in Table 4.4. Table 4.5 shows that repeated measures ANOVAs (two-tailed) revealed a significant improvement in performance across sub-blocks for critical errors and criterion fails in the inhibition block, and for misses, false positives, and criterion fails in the non-inhibition block (all $p \leq .02$).

Table 4.4

Mean performance scores in each sub-block

	Sub-block 1		Sub-block 2		Sub-block 3	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
			Inhibition block			
Critical error	.60	.43	.28	.33	.28	.33
Hand error	.03	.06	.04	.06	.04	.06
Miss	.21	.14	.13	.15	.17	.15
False positive	.02	.05	.02	.05	.03	.05
Criterion fail	.55	.25	.34	.27	.36	.29
without CEs	.30	.19	.18	.23	.18	.23
Response time	855	109	851	106	843	93
			Non-inhibition block			
Hand error	.05	.09	.03	.06	.03	.06
Miss	.20	.21	.12	.14	.09	.12
False positive	.04	.07	.01	.03	.01	.02
Criterion fail	.24	.27	.11	.17	.10	.17
Response time	835	121	806	103	816	91

Table 4.5

Repeated measures ANOVAs comparing mean performance scores across sub-blocks

Measure	*df*	*F*	*p*	η_p^2
		Inhibition block		
Critical error	2, 38	10.12	< .001	.35
Hand error	2, 38	.23	.79	.01
Miss	2, 38	2.91	.07	.13
False positive [a]	1.43, 27.19	.17	.77	.009
Criterion fail	2, 38	7.45	.002	.28
without CEs	2, 38	3.17	.06	.14
Response time	2, 38	.22	.80	.01
		Non-inhibition block		
Hand error	2, 38	.89	.42	.05
Miss [a]	1.43, 27.08	4.98	.02	.21
False positive [a]	1.43, 28.00	4.62	.03	.20
Criterion fail	2, 38	4.31	.02	.19
Response time	2, 38	1.84	.17	.09

Note. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

Consistent with previous experiments, satisfaction of the requirement to inhibit a prepotent response to critical items did not prevent subsequent critical errors for 65% of participants, despite a significant reduction in the mean number of critical errors produced with practice. Critical errors across trials for three participants with differing Culture Fair scores as well as the percentage of participants committing a critical error on each trial is presented in Table 4.6. There was a general decrease in the percentage of participants making a critical error from trial 1 to 10, but this percentage rose again for the final two trials. Again, participants often immediately noticed that they had made these mistakes.

Table 4.6

Pattern of critical error across trials for three participants with differing Culture Fair scores and percentage of participants committing critical error in each trial

Trial	Pattern of critical error			% of participants
	Participant A (3 Culture Fair errors)	Participant B (8 Culture Fair errors)	Participant C (18 Culture Fair errors)	
1	0	1	1	65
2	0	0	1	55
3	1	1	1	60
5	0	0	1	30
7	0	0	0	35
8	0	0	0	15
10	0	0	0	15
11	0	1	0	30
12	0	0	1	40
Overall	1	3	5	65

g correlations. Pearson's correlations (one-tailed) between sub-block performance scores and Culture Fair error are presented in Table 4.7. In the inhibition block, Culture Fair error was significantly correlated with misses and criterion fails (without CEs) in every sub-block, response time in every sub-block, and critical errors, false positives, and criterion fails (with CEs) in sub-blocks 2 and 3. In the non-inhibition block, significant correlations with Culture Fair error were restricted to response time in every sub-block, false positives in sub-block 1, and hand errors in sub-block 3. Williams-Hotelling t-test (two-tailed) showed that, in the inhibition block, the correlation with Culture Fair error for response time was significantly stronger in sub-block 2 relative to sub-block 3, $t(17) = 2.40$, $p = .03$. In the non-inhibition block, the hand error–Culture Fair error correlation was significantly stronger in sub-block 3 relative to sub-block 1, $t(17) = 2.38$, $p = .03$; the false positive–Culture Fair error correlation was significantly stronger in sub-block 1 relative to sub-block 2 of the non-inhibition block, $t(17) = 2.32$, $p = .03$.

Table 4.7

Pearson's correlations between Culture Fair error and sub-block performance

	Sub-block 1		Sub-block 2		Sub-block 3	
Measure	*r*(18)	*p*	*r*(18)	*p*	*r*(18)	*p*
			Inhibition block			
Critical error	.33	.08	.47	.02	.44	.03
Hand error	.06	.41	.13	.30	.01	.48
Miss	.47	.02	.62	.002	.54	.007
False positive	.18	.22	.58	.003	.67	.001
Criterion fail	.32	.08	.67	.001	.62	.002
without CEs	.60	.003	.62	.002	.47	.02
Response time	.52	.01	.64	.001	.37	.05
			Non-inhibition block			
Hand error	-.12	.31	.13	.30	.51	.01
Miss	.25	.15	.21	.19	.13	.29
False positive	.43	.03	-.18	.22	.10	.34
Criterion fail	.22	.18	.28	.12	.10	.34
Response time	.57	.004	.57	.005	.39	.04

Comparisons across rule groups. The performance scores and correlations with Culture Fair error observed in Experiment 3 were compared with those observed in Experiments 1 and 2. Independent samples t-tests (two-tailed) showed that error was significantly greater in Experiment 3 (four-rule-ONLY) relative to Experiment 2 (four-rule) for false positives in sub-block 1[14], $t(58.99) = 2.43$, $p = .02$, $d = .62$. Fisher's z-tests (two-tailed) revealed that correlations with Culture Fair error were significantly higher in Experiment 3 relative to Experiment 1 (two-rule) for the following inhibition block measures: response time in sub-block 1, $z(68) = 2.43$, $p = .02$; criterion fails in sub-block 2, $z(68) = 2.33$, $p = .02$; response time in sub-block 2, $z(68) = 2.64$, $p = .01$; and false positives in sub-block 3, $z(68) = 2.33$, $p = .02$. A marginal effect, in the same direction, was observed for critical errors in sub-block 2, $z(68) = 1.94$, $p = .052$. The correlation with Culture Fair error for response time in sub-block 2 of the non-inhibition block was also significantly higher in Experiment 3 than in Experiment 1, $z(68) = 2.08$,

[14] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

p = .04. Correlations with Culture Fair error were significantly stronger in Experiment 3 relative to Experiment 2 for: false positives in sub-blocks 2 (inhibition block), $z(68)$ = 2.13, p = .03; response time in sub-block 2 (inhibition block), $z(68)$ = 2.93, p = .003; false positives in sub-blocks 3 (inhibition block), $z(68)$ = 2.51, p = .01; response time in sub-block 1 (non-inhibition block), $z(68)$ = 2.39, p = .02; and response time in sub-block 2 (non-inhibition block), $z(68)$ = 2.22, p = .03.

Order effects. ***Performance.*** Performance scores are compared across order groups in Table 4.8. Independent samples t-tests (two-tailed) showed that inhibition block error was significantly greater in the non-inhibition–inhibition group (i.e., participants that completed this task first) relative to the inhibition–non-inhibition group for all measures other than criterion fails (p = .005 to .04). Culture Fair scores were equivalent across order groups.

Table 4.8

Mean performance scores across order groups

	Order group				Independent t-test			Cohen's
	Inhibition–non-inhibition		Non-inhibition–inhibition					
Measure	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	*d*
Culture Fair	7.90	5.49	10.30	6.18	18	.92	.37	-.41
			Inhibition block					
Critical error	.37	.26	.41	.34	18	-.32	.75	-.13
Hand error	.03	.04	.04	.03	18	-.39	.70	-.29
Miss	.16	.10	.19	.15	18	-.54	.60	-.24
False positive	.03	.04	.02	.03	18	.25	.81	.29
Criterion fail	.37	.20	.46	.25	18	-.85	.41	-.40
without CEs	.19	.13	.24	.20	18	-.70	.49	-.30
Response time [a]	834	59	864	119	13.19	-.70	.49	-.33
			Non-inhibition block					
Hand error	.01	.03	.07	.05	18	-3.05	.007	-1.50
Miss	.08	.10	.20	.14	18	-2.22	.04	-1.00
False positive [a]	.005	.007	.03	.03	9.86	-2.59	.03	-1.35
Criterion fail	.09	.16	.21	.16	18	-1.65	.12	-.75
Response time	792	105	923	75	18	-3.21	.005	-1.46

Note. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

g correlations. Pearson's correlations (one-tailed) between performance and Culture Fair error across order groups are displayed in Table 4.9. Correlations with Culture Fair error appeared larger in the non-inhibition–inhibition group for the majority of measures, but Fisher's z-test (two-tailed) revealed that none of the correlations differed significantly between order groups (all $p > .10$).

Table 4.9

Pearson's correlations between Culture Fair error and order group performance

	Order group			
	Inhibition–non-inhibition		Non-inhibition–inhibition	
Measure	*r*(10)	*p*	*r*(10)	*p*
		Inhibition block		
Critical error	.43	.11	.53	.06
Hand error	-.003	.50	.18	.31
Miss	.52	.06	.72	.009
False positive	.59	.04	.72	.006
Criterion fail	.55	< .05	.71	.01
without CE	.71	.01	.79	.004
Response time	.44	.1	.65	.02
		Non-inhibition block		
Hand error	.20	.29	.08	.41
Miss	-.08	.43	.34	.17
False positive	.22	.28	.29	.21
Criterion fail	.08	.42	.29	.21
Response time	.12	.38	.65	.02

Note. $n = 10$ in each order group.

Comparisons across rule groups. Performance scores and Culture Fair correlations in each order group were compared with those observed in the order groups of Experiments 1 and 2. Independent t-tests (two-tailed) showed that, for non-inhibition–inhibition participants, response time (in the non-inhibition block) was significantly longer in: the four-rule-ONLY group relative to the two-rule group, $t(33) = -3.38$, $p = .002$, $d = -1.35$; and in the four-rule-ONLY group relative to the four-rule group, $t(33) = -4.32$, $p < .001$, $d = -1.71$. For inhibition–non-inhibition participants, error was greater (in the inhibition block) in the four-rule group relative to the four-rule-ONLY group for: criterion fails, $t(33) = 3.25$, $p = .003$, $d = .38$; criterion fails without CEs, $t(33) = 2.29$, $p = .03$, $d = .20$; and misses, $t(33) = 2.11$, $p = .04$, $d = .85$.

Fisher's z-tests (two-tailed) revealed that the correlation with Culture Fair error was significantly higher in the four-rule-ONLY group relative to the two-rule group for a

number of measures: false positives and criterion fails (without CEs), in the inhibition block, in inhibition–non-inhibition participants; misses, false positives, criterion fails (with and without CEs) in the inhibition block, and response time in both blocks, in non-inhibition–inhibition participants (all $p < .02$). The correlation with Culture Fair error was significantly stronger in the four-rule-ONLY group, relative to the four-rule group, for a number of measures in non-inhibition–inhibition participants (misses and false positives in the inhibition block, and response time in both blocks); the opposite pattern of stronger correlations in the four-rule group was observed for hand errors in both blocks (all $p < .02$).

Learning and reconceptualisation of task rules. After task execution, participants were asked to state task rules as they were represented in their minds. Participants were categorised as reconceptualising task rules if they stated fewer than four rules[15], which was assumed to reflect reconceptualisation of rule information into a smaller number of chunks than was imposed by initial task instructions. Independent t-tests (two-tailed) showed that participants that reconceptualised the rules produced fewer Culture Fair errors ($n = 13$, $M = 6.54$, $SD = 3.78$) than participants that did not reconceptualise the rules ($n = 7$, $M = 13.86$, $SD = 6.18$) when they were instructed to state the rules as they were represented in their own minds (i.e., after the second block), $t(19) = -3.31$, $p = .004$, $d = -1.47$. However, Culture Fair error was statistically equivalent across people that did, and people that did not, reconceptualise when they were simply asked to repeat the rules (i.e., after the first block; $p = .30$).

Performance scores were compared across participants that reconceptualised task rules and participants that did not using ANCOVAs (two-tailed) in order to control for the significant difference in Culture Fair scores between groups. Participants that reconceptualised task rules (*adjusted M* = .005, *SD* = .04), relative to participants that did not (*adjusted M* = .04, *SD* = .009), made significantly fewer false positives in the non-inhibition block, $F(1, 17) = 10.01$, $p = .006$, $\eta_p^2 = .37$; the same effect was found for

[15] All participants that reconceptualised the rules stated *fewer* than four rules.

criterion fails (without CEs) in the inhibition block (reconceptualisers: *adjusted M* = .15, *SD* = .07; non-reconceptualisers: *adjusted M* = .33, *SD* = .22), $F(1, 17) = 11.94$, $p = .001$, $\eta_p^2 = .58$.

Rule failure scores and total proportion error scores. Participants were assigned rule failure and proportion error scores for each block (see Chapter 2, page 60 for a description of how these scores were calculated). Paired t-tests (two-tailed) revealed that rule failure scores were significantly higher in the inhibition block relative to the non-inhibition block, $t(19) = 3.24$, $p = .004$, $d = 1.06$. Proportion error scores were also significantly higher in the inhibition block relative to the non-inhibition block, $t(19) = 3.89$, $p = .001$, $d = 1.10$. However, Williams-Hotelling t-test (two-tailed) showed that correlations with Culture Fair error for proportion error scores were statistically equivalent across blocks, $t(17) = 1.73$, $p = .10$. A marginal effect was observed for rule failure scores, with a stronger correlation in the inhibition block, $t(17) = 1.99$, $p = .06$.

Table 4.10

Performance and Pearson's correlations with Culture Fair for rule failure and proportion error scores

	Inhibition block				Non-inhibition block			
	Performance		*g* correlation		Performance		*g* correlation	
Score	*M*	*SD*	*r*(18)	*p*	*M*	*SD*	*r*(18)	*p*
Rule failure	.60	.82	.38	< .05	.05	.22	-.09	.36
Proportion error	.21	.12	.71	< .001	.10	.08	.32	.08

ANCOVAs (two-tailed) revealed that neither rule failure scores nor total proportion error scores differed significantly from those observed in Experiment 1 or Experiment 2 in the inhibition block, the non-inhibition block, or both blocks combined (all $p > .10$). Fisher's z-tests (two-tailed) revealed that the correlation with Culture Fair error for rule failure scores also did not differ statistically across rule groups for either block or both blocks combined (all $p > .10$). The correlation with Culture Fair error for proportion error scores was significantly stronger in Experiment 3 relative to Experiment 1 for the

inhibition block, $z(68) = 1.96$, $p = .05$; a marginal effect in the same direction was observed for both blocks combined[16], $z(68) = 1.71$, $p = .09$.

Real-time task execution demand versus task conceptualisation complexity. Task conceptualisation (or the number of rules in the participants' minds, ranging from one to four) was significantly correlated with Culture Fair error at $r_s(18) = .56$, $p = .01$. Spearman's rank order correlations between task conceptualisation and performance scores are presented in Table 4.11. Task conceptualisation was significantly correlated with false positives and criterion fails (without CE) in the inhibition block. In order to determine whether or not real-time task execution demand was driving the correlation between task conceptualisation and Culture Fair error, a series of partial correlations were performed; these partial correlations between task conceptualisation and Culture Fair error separately controlled for the score on each performance measure. Similarly, partial correlations were performed between performance scores and Culture Fair error, each controlling for task conceptualisation, in order to assess any mediating effect of the number of task rules held in mind. Fisher's z tests (two-tailed) revealed that none of these correlations differed significantly from their bivariate equivalents (all $p > .10$).

Table 4.11

Spearman's correlations between task conceptualisation and performance

	Inhibition block		Non-inhibition block	
Measure	$r_s(18)$	p	$r_s(18)$	p
Critical error	.14	.56	-	-
Hand error	.17	.49	.31	.19
Miss	.38	.10	-.03	.89
False positive	.80	< .001	.40	.08
Criterion fail	.28	.23	-.02	.94
without CEs	.46	.04	-	-
Response time	.41	.07	-.01	.98

[16] The correlation with Culture Fair error for proportion error scores in the four-rule-ONLY group was $r(18) = .69$, $p < .001$.

Discussion

Performance scores were consistent across Experiments 1 through 3 for the majority of measures. However, the involvement of *g* in parts of the task that required response inhibition was statistically significant only in the two groups that received task instructions as four rules (Experiments 2 and 3), and not the group that received task instructions as two rules (Experiment 1). Across all three experiments, critical items that required inhibition of a prepotent but inappropriate response tendency were rather problematic for participants as suggested by: (a) greater error rate on critical items relative to other measures (with the exception of criterion fails that took into account performance across task elements); (b) greater error in the block that contained critical items relative to the block that did not for criterion fails (in all three experiments), misses, and response time (each in Experiments 1 and 2 only); and (c) success on critical items did not prevent later critical error for a large proportion of the sample (65% in Experiment 3; 78–80% in previous experiments). The only measures for which performance differed across rule groups were response time (in the non-inhibition block) and false positives (in sub-block 1 of the inhibition block); response time was shorter, and false positives to non-matching frames were more frequent, in the four-rule-ONLY group relative to the four-rule group.

Collectively, the correlational findings show that the relationship between response inhibition (performance on critical items) and *g* (performance on the Culture Fair) depends on the number of rules presented at task instructions, regardless of whether there is an added requirement for reconceptualising the task when learning task rules (Experiment 2) or not (Experiment 3). This is suggested by two observations. First, critical errors were significantly correlated with *g* in Experiments 2 and 3 (i.e., in the four-rule groups), but not in Experiment 1 (i.e., in the two-rule group). Second, only in Experiments 2 and 3 was the correlation between *g* and some performance measures significantly stronger in the block that contained critical items relative to the block that did not.

Direct comparisons of correlations across experiments were not so clearly supportive of the overarching importance, to the recruitment of *g*, of the number of rules presented at task instructions relative to a requirement to reconceptualise the task during the presentation of task rules. Experiment 2 showed that the correlations with Culture Fair scores for response inhibition measures (critical errors and criterion fails in the inhibition block) were significantly stronger in Experiment 2 (when the task was conceptualised as four task rules that required reconceptualisation) relative to Experiment 1 (when the task was conceptualised as two task rules that did not require reconceptualisation). This suggested that an increase in the number of task rules, coupled with an added requirement to reconceptualise the task whilst task rules are being learned, increases the involvement of *g* in task performance. Separating out these factors in Experiment 3, however, resulted in the emergence of results that were quite difficult to interpret.

The investigations reported in Chapter 4 revealed statistically equivalent correlations between *g* and response inhibition measures across Experiments 2 and 3 (i.e., between the two groups that each received four rules, but which differed in a requirement to reconceptualise task rules). This suggests that the presence of a requirement to reconceptualise the task, whilst learning task rules, does not act to increase task-sensitivity to *g* when the number of presented task rules is held constant. However, this view is complicated by the observation of statistically equivalent correlations between *g* and response inhibition measures across Experiments 1 and 3 also (i.e., between the two groups that were each not required to reconceptualise task rules, but which received a different number of rules). This suggests that an increase in the number of task rules does not act to increase task-sensitivity to *g* when demand for reconceptualisation of the task model, as it is being formed, is held constant.

For other measures (namely false positives and response time), the correlation with *g* was significantly stronger in Experiment 3 relative to each of the previous experiments (in which these correlations were negligible). However, the correlation

with *g* was not significantly (and also not always numerically) stronger for critical errors relative to measures that did not involve inhibition in Experiments 1 and 3 (unlike Experiment 2). In Experiment 3, the correlation between *g* and response time was more like what would be expected by the processing speed explanation of intelligence (e.g., Grudnik & Kranzler, 2001). The full body of results so far, however, show somewhat inconsistent relationships between response speed and *g*. There is no obvious explanation for the emergence of significant correlations between response time and *g* in Experiment 3. It is unlikely that some kind of speed/accuracy trade-off that may be important to *g* was involved in some experiments but not others. This is because the experiments thus far differed only in either the number of rules presented to participants or the requirement to reconceptualise task rules on formation of the task model. In other words, the experiments did not differ in terms of the operative task-relevant information presented to participants, nor did they differ in terms of the actual requirements of the task. In any case, it was difficult to explore whether, for example, lower *g* participants responded quickly but inaccurately (or vice versa) because response time scores were only collated for correct responses.

The correlational findings go some way in supporting the theory that, in conditions of response inhibition demand only, the number of rules affects the level of involvement of *g* in task performance, over and above any effects of the added requirement to reconceptualise the task when learning task rules. However, the sample size in Experiment 3 was small ($N = 20$), particularly for correlational research. Although smaller samples generally provide less stable correlations, this sample is not sizably smaller than those of relevant published research in the field (e.g., $N = 24$, Dumontheil et al., 2010). The co-efficient for the correlation between critical errors and *g* was numerically greater in Experiment 3 ($r = .49$) relative to Experiment 1 ($r = .20$), and was numerically similar across Experiments 2 ($r = .54$) and 3. Thus, it is possible that if a larger sample was recruited for Experiment 3, that the critical error–*g* correlation would have contained enough variability to be *significantly* stronger than that observed in

Experiment 1 (but remain unchanged from that observed in Experiment 2). Because this is by no means guaranteed, the conclusion that the number of rules is more important to the recruitment of *g* than a requirement to reconceptualise the task whilst task rules are being learnt remains tentative.

Another issue with regards to task reconceptualisation that was addressed by Experiment 3 was whether some participants, when encouraged to think about how the task is represented in their own mind, would state a quite different conceptualisation of the task to that encouraged by the four rules set out in task instructions. The findings showed that every participant that showed evidence of reconceptualisation (using the defined boundaries) effectively reduced the number of distinct rules comprised in their task conceptualisation. Furthermore, participants that reconceptualised the task produced significantly fewer Culture Fair errors relative to participants that did not reconceptualise the task, but only when they were explicitly asked to impose their own order on instructions (and not when they were asked to simply repeat the rules). Thus, when encouraged to state the rules as they were conceptualised in mind, participants that scored higher on the Culture Fair test reconceptualised the task, stating a less complex conceptualisation that involved fewer distinct rules. This provides direct evidence for the theory that has been suggested by the correlational findings – that *g* may reflect the ability to apply a relatively efficient task model to behaviour.

The importance of task modelling in task-recruitment of *g* may provide some explanation for the phenomenon of positive manifold observed between scores on cognitive tasks (e.g., Carroll, 1993). Perhaps all complex tasks, irrespective of type, require an efficient task model in order to be completed effectively. For relatively less complex, and relatively more automatic, tasks, efficient task conceptualisation may not be as critical for correct performance. Indeed, Experiments 1 through 3 have shown that the number of rules presented in task instructions affects the recruitment of *g* in the arguably more complex block (i.e., the block that contained critical items), but not in the arguably less complex block (i.e., the block that did not contain critical items).

Although performance did not differ between groups that received a different number of rules at the time of task instructions, participants that reconceptualised task rules, forming a more efficient task model, also performed better on some measures, independent of where they fell on the *g* distribution. This provides some support that efficient task modelling is central for correct performance of complex tasks.

A close relationship between *g* and individual differences in the ability to effectively conceptualise task rules could also explain the capacity of *g* to predict a wide range of behaviours (e.g., Spearman, 1904, 1927). Fewer task rules could allow for a greater availability of working memory or attentional resources for real-time task demands (because these resources are not "used up" by the storage of task-relevant information), rendering performance of the task easier. That performance did not differ between two- and four-rule groups does not necessarily contradict this possibility. Although a basic assumption of the present research was that the format of presented instructions corresponds to the way in which this information is mentally represented, this may not be the case for all individuals (particularly for those participants, with higher *g*, that reconceptualised task rules). An avenue for subsequent experiments (Experiment 6) involves asking participants to impose their own order on task instructions *prior* to task execution. It was hoped that this would provide a more realistic measure of the specific conceptualisations[17] that are applied to task performance, and would therefore shed light on how this affects performance levels and the recruitment of *g*.

The characteristics of task conceptualisation appear to be of greater relevance to Spearman's *g* than the characteristics of the performed task per se. This was suggested by the observation of (a) significant response inhibition–*g* correlations, and (b) significant differences in performance–*g* correlations between the block that demanded response inhibition and the block that did not, only when task instructions were presented as four distinct rules. However, direct examination of this theory yielded no

[17] Whether participants reconceptualise the task pre-, during, or post-task performance is extremely difficult to measure, and this complicates conclusions.

significant findings. The number of rules represented in the participants' own task conceptualisations (which ranged from one to four) was significantly correlated with *g*, and the strength of this correlation did not change significantly when partial correlations were computed that controlled for each performance measure in turn. Furthermore, although a number of task measures (particularly those measures in the inhibition block) were significantly correlated with *g*, the strength of these correlations did not change when partial correlations were computed that controlled for the number of rules comprised in the participants' task conceptualisations. Experiments 4, 5, and 6 explored these same issues using a more precise measure of task conceptualisation complexity. This involved participants selecting a task conceptualisation that best reflected their own task conceptualisation from a rule sheet; this rule sheet offered a selection of task conceptualisations which differed solely in the number of rules that they comprised.

In conclusion, Experiment 3 provided some evidence that the number of rules contained in the conceptualisation of the task, rather than a requirement to alter the conceptualisation of the task formed on the basis of early-specified rules, appears to be the important aspect of four-rule instructions that increases the likelihood of the recruitment of *g*. Response inhibition, whilst possibly a fundamental risk factor for the recruitment of *g*, was only so in the present experiments when task instructions were presented as four rules (and not when instructions were presented as two rules). *g* may therefore reflect the ability to form an "efficient" conceptualisation of the task at hand; this conceptualisation may be used to control task performance and may be especially important when tasks are high in complexity (e.g., when demand on response inhibition is present). In Experiments 4 and 5, task conceptualisation complexity was manipulated in the same way as in Experiments 1 through 3 (i.e., by splitting task requirements into two or four rules), but a different experimental task was performed. This new task placed varying levels of demand on a number of cognitive processes postulated as being involved in *g*—sustained attention or maintenance, updating, and inhibition—in order to

investigate whether or not the findings extend beyond the constraints of the colour shape match task.

CHAPTER 5

Real-time Task Demand and Spearman's *g*

Overview

Previous experiments have indicated that relatively inefficient modelling of task goals increases task-sensitivity to Spearman's *g*, particularly in conditions that require the inhibition of a prepotent response tendency. Chapter 5 presents a new task—the Dot Pattern Expectancy (DPX) task—to establish whether these observations extend to types of real-time cognitive demand other than response inhibition. Specifically, in Experiment 4, the level of demand imposed by the maintenance and updating of information and two types of inhibition (involving an alternative response, rather than no response, and including inhibition of a *prepared* response), was systematically manipulated across task performance. Task instructions were presented as two distinct rules. The results revealed that, in the context of this efficient task conceptualisation, very few performance measures were significantly correlated with *g*, and manipulations of real-time execution demand did not affect task-sensitivity to *g*. Furthermore, the results confirmed, using a more objective measure of task reconceptualisation, that efficient reconceptualisation of task goals is restricted to people at the higher end of the *g* distribution. It is concluded that real-time task execution demands are only weakly associated with Spearman's *g* in the context of an explicitly defined efficient task model.

Experiment 4

The findings of Experiments 1 through 3 show that as the number of distinct chunks, or rules, presented in task instructions increases, so does the strength of correlations between task performance and *g*. This is quite independent of task difficulty: first, because the number of task requirements was held constant across all

three rule conditions (the participants in Experiments 1, 2, and 3 carried out the same task and received the same task information); and second, because the performance scores across all three rule groups were statistically equivalent. However, it was only in the context of a requirement to inhibit a prepotent response tendency that an increase in the number of presented task rules strengthened correlations with *g*. These findings suggest that, at the cognitive level, Spearman's *g* might be considered as some kind of control function that is associated with the efficient internal representation of (novel) task-relevant information, particularly in tasks that require response inhibition. A logical next direction for investigation is to consider whether these effects are unique to response inhibition. Indeed, it is possible that the *level*, rather than the *type*, of processing demand results in strengthened *g* correlations via increasing the number of presented rules.

Experiment 4 explores these issues by introducing a new task, the DPX task (MacDonald et al., 2005). The DPX task is a modification of the classic AX continuous performance task (AX-CPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956) and is the non-verbal equivalent of the Expectancy AX task (Servan-Schreiber, Cohen, & Steingard, 1996). In the Expectancy AX task, letters are presented as cue-probe pairs with a delay in between the presentation of the cue and the presentation of the probe; the delay varies in duration such that sometimes it is short in duration (1,000 ms, e.g., McClure et al., 2010; Barch et al., 2004; MacDonald et al., 2005) and sometimes it is long in duration (5,000 ms, e.g., Barch et al., 2004; McClure et al., 2010; 8,000 ms, e.g., MacDonald et al., 2005; or 9,500 ms, e.g., Carter et al., 1998). Participants respond by pressing one of two buttons: a target key or a non-target key. At the time of probe presentation, the target key should be pressed when the cue-probe pair is AX (if the probe is X and the preceding cue was A), but the non-target key should be pressed when the cue-probe pair is not AX. There are three types of non-target pair that are used for the purpose of scoring: AY pairs which comprise an A cue but a non-X probe; BX pairs which comprise a non-A cue but an X probe; and BY pairs which correspond to a non-

A cue and a non-X probe. To assist encoding of each cue, the non-target key is also pressed at the time of cue presentation, regardless of whether the cue is valid (A) or invalid (B).

The Expectancy AX task is often used as a measure of context processing deficit (e.g., Cohen, Barch, Carter, & Servan-Schreiber, 1999; MacDonald et al., 2005). MacDonald et al. (2005) describe context processing as "the component of cognitive control that represents and actively maintains task-relevant information despite subsequent noise" (p. 814), and link it to selective attention (Norman & Shallice, 1986), distractibility (Oltmanns & Neale, 1975), and executive processing (Baddeley & Hitch, 1974). Each cue in the Expectancy AX task provides the context for preparing a response to the probe, and this context needs to be internally represented and maintained across the delay between cue and probe presentations. A cues provide the context for producing a target response to an X probe, whereas B cues provide the context for producing a non-target response to any probe (X or Y). A high percentage (70%) of trials in the Expectancy AX task correspond to AX cue-probe pairs, which results in target responses being facilitated, or prepared, by the presentation of A cues. Therefore, a different pattern of error is observed depending on whether or not context processing is "intact."

Intact context processing results in frequent AY error but infrequent BX error; if the cue is successfully maintained across the delay then, in AY trials, the target response that is prepared by the A cue needs to be inhibited on presentation of the Y probe, but in BX trials, the non-target response that is prepared by the B cue prevents the presentation of the prepotent target X probe from overturning the appropriate non-target response. Impaired context processing, however, in which the cue is not successfully maintained, results in the opposite pattern of error because responses are made solely on the basis of the probe; in AY trials, a target response is not prepared by the A cue so the non-target response indicated by the Y probe is correct, and in BX trials, a non-target response is not prepared by the B cue so the prepotent target response indicated by the X probe is

incorrect. Furthermore, in intact context processing, the facilitated response is strengthened by a longer delay such that AY trials are more error-prone and BX trials are less error-prone in long, relative to short, delay conditions (MacDonald et al., 2005).

The Expectancy AX task is appropriate for use in this research project for two main reasons. First, MacDonald et al.'s (2005) definition of context processing—the ability to represent, maintain, and utilise task-relevant, contextual information in order to guide behaviour—is comparable to Duncan et al.'s (2008) description of the task model as "a working memory description of relevant facts, rules, and requirements used to control behaviour" (p. 140). On these grounds, if *g* reflects (or is at least associated with) some kind of capacity for maintaining, re-modelling, and utilising task-relevant information (see also Kane & Engle, 2002) then correlating performance on the Expectancy AX task with *g* may provide further insight into the cognitive basis of *g*. Significant correlations have indeed been observed between context processing and *g* in healthy populations (MacDonald et al., 2005; discussed in more detail below), but how this relationship is affected by the presentation of task rules has not been addressed in the literature. Second, the Expectancy AX task affords the separation and manipulation of a number of different task demands: inhibition of the prepared target response indicated by the cue on AY trials; inhibition of the prepotent target response indicated by the probe on BX trials; updating of contextual information (when the cue to be maintained changes, for example, from A to B); and maintenance of contextual information (across a variable delay).

There are, however, a number of drawbacks associated with the use of the verbal version of the task in the current project. First, the Expectancy AX task usually takes a long time to complete (> 20 min) because, in order to build up a prepotency to the two types of response bias (i.e., a prepared target response following A cues and a target response to X probes), it is important that AX trials are high in abundance. Second, ceiling effects are sometimes observed in AY trials in healthy samples (e.g., Cohen et al., 1999). Third, the verbal nature of the stimuli may be inappropriate for these

investigations because it is possible that some participants make associations between the target stimuli and, for example, think of the word or object *axe* and use this cue to initiate a target response. MacDonald et al. (2005) argue that the severity of these disadvantages is reduced in the visuospatial version of the task.

In the DPX task, letters are replaced by dot pattern configurations similar to Braille font (see Table 5.1, page 124). MacDonald et al. (2005) designed the task such that non-target B cues are distinctive from the target A cue, whereas non-target Y probes are quite similar to the target X probe. They did this in order to (a) reduce the likelihood of ceiling effects on AY trials (because Ys are more similar to Xs), and (b) reduce the number of trials that need to be completed (because AY errors accumulate faster). MacDonald et al. (2005) also argue that because novel visuospatial stimuli degrade faster than common overlearned stimuli such as letters, the delay between cue and probe presentations can be shorter whilst eliciting the same behavioural effects. Another benefit of employing the DPX task is that greater estimation power is provided for the critical trial types because the frequency of BX and AY trials are increased to 12.5% each (relative to 10% in the Expectancy AX task). Through confirmatory factor analysis MacDonald et al. (2005) found that both the AX and the DPX tasks tap the same two uncorrelated processes—context processing, which is indexed by AX and BX performance, and response preparation, which is indexed by AY trials and, to a lesser extent, AX trials—despite the adoption of stimuli from different domains. The authors therefore suggest that context processing is modality neutral; the cue is not stored as A or B but rather as a "representation associated with its significance ('likely to press target' or 'must not press target, press nontarget instead')" (p. 819).

As well as developing and assessing the validity of the DPX task, MacDonald et al. (2005) investigated context processing deficits in schizophrenia. Central to this research project, they also explored the relationship between Expectancy AX and DPX task performance, general intelligence, and working memory in a large sample of healthy adults. For their healthy controls, simple methodological differences

significantly affected performance, which is a key consideration concerning the use of this task for the current study. For example, the trend for controls to produce greater error on AY trials relative to BX trials (due to intact context processing) was emphasised by the presence of experimental supervision and a reminder to respond as quickly and as accurately as possible between blocks.

In MacDonald et al.'s (2005) "best fit" model, the general intelligence and working memory factors were significantly correlated with each other at $r = .61$, which is similar to the co-efficient reported in other studies (e.g., Süb et al., 2002). Working memory was significantly correlated with both context processing ($r = .50$) and response preparation ($r = .37$) factors, whereas general intelligence was significantly correlated only with context processing ($r = .53$). These findings are consistent with the notion that general intelligence and working memory are related but independent constructs (e.g., Conway et al., 2003), otherwise both would be related to response preparation. Due to the strong relationship between each of these constructs and context processing, the authors conclude that "these more complex processes may rely on the ability to flexibly represent and maintain context to control behaviour" (p. 819). This is consistent with fMRI research which shows that context representation and maintenance recruit the dorsolateral prefrontal cortex, an area associated with traditional aspects of working memory (D'Esposito et al., 1995; Sakai, Rowe, & Passingham, 2002), general intelligence (Duncan & Owen, 2000; Gray & Thompson, 2004), and task modelling (Dumontheil et al., 2010).

Some aspects of the DPX task were used as measures of response inhibition in the current experiment. However, the characteristics of inhibition demand in the DPX task were different to those in the colour shape match task used in previous experiments. First, the level of prepotency associated with target responses was greater in the DPX task relative to the colour shape match task because target trials occurred with greater frequency; 70% of DPX trials were AX pairs whereas only 30% of colour shape match trials, in the inhibition block, were single-matching pairs. However, this greater

prepotency of target trials in the DPX task was balanced by the inclusion of a greater number of trials that demanded inhibition; in the DPX task, 12.5% of trials required inhibition of a prepared response (AY pairs) and a further 12.5% of trials required inhibition of a prepotent response to the probe (BX pairs), whereas in the colour shape match task, only 7.5% of items required inhibition (double-matching pairs). Second, inhibition trials required a different response (or non-response); double-matching pairs required the inhibition of a motor response in favour of *no* response, whereas AY and BX pairs required the inhibition of a motor response in favour of a *different* motor response. Another difference between the two tasks was that *response conflict* was not associated with single-matching pairs but it was associated with AX pairs; response conflict was associated with all trials that contained an A-cue or an X-probe because A-cues and X-probes each indicated a target only 85% (and not 100%) of the time.

In Experiment 4, the DPX task was used to assess the contributions of inhibition, updating (of contextual information), and maintenance (or context processing or sustained attention) demands to the recruitment of *g* in task performance when task rules are presented in two distinct chunks. Inhibition demand was manipulated by trial type: inhibition demand was present in AY and BX trials but was absent from AX and BY trials. Maintenance of contextual information was manipulated by the length of delay between cue and probe presentations: maintenance demand was high in the *long-delay block* but low in the *short-delay block*. The majority of studies that have employed the AX task present a continuous block of short-delay trials and a continuous block of long-delay trials because switching trial-by-trial reduces the effects of prepotency manipulations (e.g., MacDonald et al., 2005; Barch et al., 2004). Updating of contextual information may be manipulated by cue type (provided that the cue is maintained throughout the trial and does not degrade until the presentation of the following cue): updating demand was present when the current cue was different from the previous presented cue (i.e., a B-cue following an A-cue or an A-cue following a B-

cue; termed *updating trials*) but was absent when the cue remained constant (i.e., an A-cue following an A-cue or a B-cue following a B-cue; termed *non-updating trials*).

Another key feature of Experiment 4 was that participants selected a rule-format that best reflected how the rules were represented in their own mind from a sheet outlining five possible formats (which presented task instructions in one, two, three, or four rules, each comprising the same amount of operative task-relevant information)[18]. Participants were encouraged to write their own description of task rules if none of the options were a good reflection of how task rules were represented in mind. It was hoped that this would provide a more objective measurement of task reconceptualisation than an analysis of participants' verbal description of rules (in which the number of rules was determined by the researcher).

On the basis of the findings of Experiments 1—that the recruitment of *g* was unaffected by manipulations of response inhibition demand in the context of the presentation of two distinct task rules—it was predicted that the correlation between task performance and Culture Fair error in Experiment 4 would be unaffected by manipulations of real-time execution demand. Specifically, correlations with Culture Fair error were expected to be statistically equivalent across (a) trial type (AX vs. AY vs. BX vs. BY), (b) block (long-delay vs. short-delay), and (c) cue type (updating vs. non-updating). Nonetheless, performance was predicted to be significantly poorer on trials, blocks, and cues that were associated with high real-time execution demand relative to those that were associated with low real-time execution demand. On the assumption that task conceptualisation efficiency is related to *g*, it was also predicted that the number of chunks of task rule information comprised in the selected rule-formats (which was assumed to reflect the number of chunks in which task requirements were represented in memory) would be significantly correlated with Culture Fair error.

[18] Essentially, these rule formats reflected two-rule instructions either merged into either one chunk (Rule 1 plus Rule 2), two chunks (Rule 1, Rule 2), three chunks (Rule 1, Rule 2 split into two parts), or four chunks (Rule 1 split into two parts, Rule 2 split into two parts).

Method

Participants

Twenty-four volunteers (17 female) aged between 18 and 53 years (*M* = 23.38 years, *SD* = 7.48) with no history of neuropsychological disorder were recruited from an online student recruitment system at Anglia Ruskin University (Psychology Department) and the wider community. None of the participants were recruited for previous experiments. Culture Fair scores in this sample ranged between 2 and 21 errors (*M* = 9.79, *SD* = 4.35).

Materials

Dot Pattern Expectancy (DPX) task. The DPX task was programmed in EPrime 2.0 (Schneider, Eschman, & Zuccolotto, 2002) and was run on a Dell PC. Stimulus frames were dot pattern configurations (100 mm × 100 mm Braille font) presented in the centre of a high resolution (1440 × 900 pixels) colour monitor. There were three to four dots in each dot pattern configuration; dots were 2 mm × 2 mm, distanced 5 mm to 1 cm apart. The dots were white (RGB: 255, 255, 255) and the background colour was grey (RGB: 169, 169, 169). Each trial consisted of two dot pattern configurations as a cue-probe pair; in each trial, an opening screen displaying a fixation cross was presented in the centre of the monitor, followed by a dot pattern configuration representing the cue A or B (with B corresponding to any configuration other than A), followed by a blank-screen delay, followed by a dot pattern configuration representing the probe X or Y (with Y corresponding to any configuration other than X), followed by a blank-screen inter-trial interval. There were two blocks of 40 trials; in the short-delay block, the blank-screen delay was short (1,000 ms) in duration and the inter-trial interval was long (5,000 ms), and in the long-delay block the delay was long (5,000 ms) in duration and the inter-trial interval was short (1,000 ms). In each block, there were 28 AX trials, five AY trials, five BX trials, and two BY trials. Trials were presented in a pseudo-random order in order to build up a prepotency to target (AX) pairs and to systematically

manipulate the levels of inhibition and updating demand. Ten trials required inhibition (five AY and five BX) and 12 trials required updating (six A cues and six B cues). Table 5.1 displays the specific dot pattern configurations used to represent A, B, X, and Y; each of the five B and Y stimuli were presented at least once in each block. Figure 5.1 displays typical long-delay trials in the DPX task.

Table 5.1

Dot pattern stimuli representing target A cues, non-target B cues, target X probes, and non-target Y probes in the DPX task

	Target	Non-targets				
Cues						
	A	B1	B2	B3	B4	B5
Probes						
	X	Y1	Y2	Y3	Y4	Y5

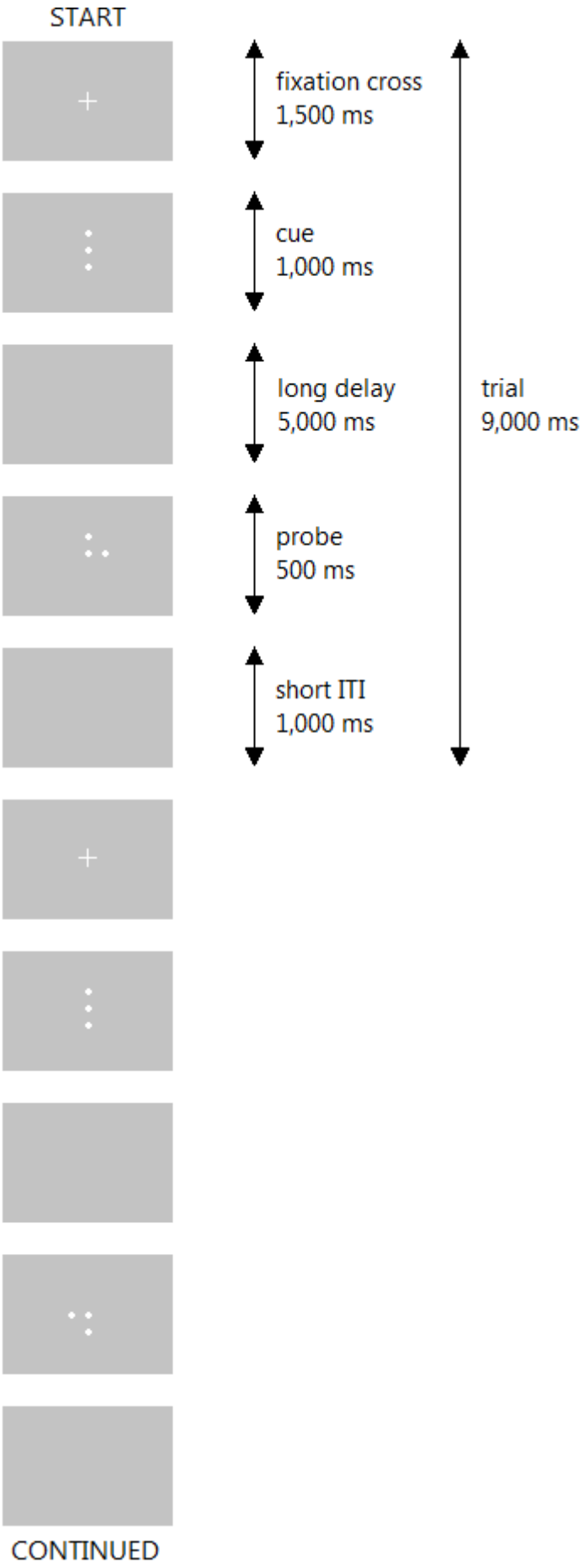


Figure 5.1. Typical long-delay trials in the DPX task.

Rule Sheet. The rule sheet (Appendix E) instructs participants to "tick the set of rules that best reflects how you thought about the task, or write your own description." There were four rule format options that each conceptualised task requirements using a different number (one, two, three, or four) of rules. Each rule format contained the same amount of operative task-relevant information. A fifth option allowed participants to write their own description of the rules.

Design

Performance measures are presented in Table 5.2. A mixed design was adopted in which performance (accuracy and response time) and correlations with Culture Fair error were compared across trial types (repeated measures: AX vs. AY vs. BX vs. BY), cue types (repeated measures: updating vs. non-updating), blocks (repeated measures: short-delay vs. long-delay), and order groups (independent samples: short–long [n = 12] vs. long–short [n = 12]). The order of blocks was counterbalanced across participants and the sequencing of trials varied pseudo-randomly; the specific sequencing of trials was maintained for all participants in order to hold the level of prepotency associated with responses to AY and BX trials constant.

Table 5.2

Definitions of performance measures for the DPX task

Trial type	Definition
AX	Target not requiring inhibition
AY	Non-target requiring inhibition of a prepotent prepared response
BX	Non-target requiring inhibition of a prepotent response
BY	Non-target not requiring inhibition
Non-updating	Cue is the same as the cue on the previous trial (A–A; B–B)
Updating	Cue is different to the cue on the previous trial (A–B; B–A)
Short-delay	Delay between cue and probe presentations is 1,000 ms (ITI = 5,000 ms)
Long-delay	Delay between cue and probe presentations is 5,000 ms (ITI = 1,000 ms)

Procedure

Participants were tested individually in a quiet and brightly lit testing room. Participants first completed the Culture Fair which was administered exactly as specified in the manual.

Participants then received instructions for the DPX task. They were informed that a series of dot patterns would be presented, one at a time, in the centre of the screen. Using an example sheet that outlined a typical run (see Appendix G), participants were instructed to view the patterns in pairs such that after the fixation cross the first pattern of the pair would be presented followed by a blank-screen delay, then the second pattern of the pair would be presented followed by another blank-screen gap before the next pair. Using another example sheet that outlined stimulus types (see Appendix H), participants were instructed to look out for target pairs, in which a target first (A) was followed by a target second (X), and that all other combinations of patterns represented non-target pairs. They were notified of the possibility that non-target pairs could share characteristics with the target pair, that is, that non-target pairs could have a target first followed by a non-target second, or a non-target first followed by a target second (or that they could differ completely from the target pair and contain two non-target patterns). Task rules were then administered. The first rule was: "when the first pattern appears, press red if it is a target or a non-target." The second rule was: "when the second pattern appears, press green if both patterns were targets and press red if either or both patterns were non-targets."

Participants then worked through an example trial with the experimenter. After this, participants were told that they would complete the task twice, once with a short delay between the presentation of the first pattern and the second pattern and once with a long delay, and were reminded that a fixation cross would indicate the start of each pair. After the rules were administered for a second time, participants were asked to repeat them. Verbal responses were recorded on a dictaphone. If the participant described the rules incorrectly (i.e., with any omissions) the appropriate rule was repeated by the

experimenter until all rules were described correctly. The monitor was situated at a distance of .50 m from the participant. The participant pressed the space bar to start the task after being instructed to respond as quickly and as accurately as possible.

In each trial, a fixation cross was presented for 1,500 ms, followed by a cue dot pattern stimulus for 1,000 ms, followed by a blank-screen delay interval of 1,000 ms (in the short-delay block) or 5,000 ms (in the long-delay block), followed by a probe dot pattern stimulus for 500 ms, followed by an inter-trial interval of 5,000 ms (in the short-delay block) or 1,000 ms (in the long-delay block). Varying inter-trial intervals in this way ensured that total trial duration was constant across short- and long-delay conditions. The durations of cue presentation, probe presentation, and short delay were identical to those used by MacDonald et al. (2005), but long delay was 1,000 ms longer in order to increase the likelihood of the recruitment of *g* (if implicated in sustained attention to task) in the long, relative to the short, delay condition. Right-handed and left-handed versions of the task were programmed so that the target key (green) was always hit with the index finger (key B in the right-handed version; key V in the left-handed version) and the non-target key (red) was always hit with the middle finger (key N in the right-handed version; key C in the left-handed version). Responses were recorded in EPrime and were attributed to a frame if they occurred within 200 ms (< 200 ms was considered anticipatory) and 1,200 ms (> 1,200 ms was considered an outlier) of stimulus (cue/probe) onset. Each block took approximately 10½ min to complete.

After the first block of 40 trials, participants were asked to repeat the rules (if any of the rules were stated in a different format to that outlined in task instructions, or if there were any omissions, the appropriate rule was repeated and the participant was asked to state the rules again). Participants were reminded to respond as quickly and as accurately as possible before the second block of 40 trials commenced. After the second block, participants were asked to think about how the rules of the task were represented *in their own mind*. Using the rule sheet, they selected the option that best

reflected how they thought about the task, or wrote their own description if none of the options reflected their own task conceptualisation. See Appendix F for the full research protocol.

Results

Main findings

Performance. Error (%) and response time (ms) scores for cue types (updating, non-updating) and trial types (inhibition: AY and BX; non-inhibition: AX and BY) across delay lengths (short-delay, long-delay) are presented in Table 5.3.

Table 5.3

Mean error and response time across cue type, trial type and delay length

	Block			
	Short-delay		Long-delay	
Measure	*M*	*SD*	*M*	*SD*
		Error (%)		
Non-updating	.30	1.01	.74	1.48
Updating	5.56	8.03	6.25	7.47
AX	3.27	4.07	2.98	3.75
AY	21.67	22.78	16.67	15.23
BX	7.50	11.52	3.33	9.63
BY	6.25	16.89	2.08	10.21
		Response time (ms)		
Non-updating	492	83	554	95
Updating	652	119	674	113
AX	522	98	519	104
AY	742	168	728	129
BX	476	176	395	113
BY	473	204	386	100

For error scores, a 4 (trial) × 2 (delay) repeated measures ANOVA (two-tailed) revealed a significant main effect of trial type[19], $F(1.86, 42.79) = 13.36$, $p < .001$, $\eta_p^2 = .37$. However there was no main effect of delay length, $F(1, 23) = 2.90$, $p = .10$, $\eta_p^2 = .11$. There was also no trial type × delay length interaction[19], $F(2.11, 48.57) = .43$, $p = .67$, $\eta_p^2 = .02$. Post hoc pairwise comparisons (two-tailed) showed that error was significantly greater on trials requiring the inhibition of a prepared response relative to all other trial types: error was greater on AY trials ($M = 19.17$, $SE = 3.29$) relative to AX trials ($M = 3.12$, $SE = .71$; $p < .001$, Bonferroni-corrected), BX trials ($M = 5.42$, SE = 1.47; $p = .004$, Bonferroni-corrected), and BY trials ($M = 4.17$, $SE = 1.94$; $p = .006$, Bonferroni-corrected). A 2 (cue) × 2 (delay) repeated measures ANOVA (two-tailed) was performed on the cue error scores which revealed a significant main effect of cue type; error was significantly greater on updating cues ($M = 5.90$, $SE = .87$) relative to non-updating cues ($M = .52$, $SE = .17$), $F(1, 23) = 40.12$, $p < .001$, $\eta_p^2 = .64$. However, there was no main effect of delay length, $F(1, 23) = .17$, $p = .68$, $\eta_p^2 = .007$. There was also no cue type × delay length interaction, $F(1, 23) = .009$, $p = .93$, $\eta_p^2 = .00$.

For response time scores, a 4 × 2 repeated measures ANOVA (two-tailed) revealed a significant main effect of trial type, $F(3, 69) = 65.97$, $p < .001$, $\eta_p^2 = .74$. There was also a significant effect of delay length, $F(1, 23) = 5.04$, $p = .04$, $\eta_p^2 = .18$. The trial type × delay length interaction was significant, $F(3, 69) = 4.70$, $p = .005$, $\eta_p^2 = .17$; this interaction is plotted in Figure 5.2 (A). Post hoc pairwise comparisons using repeated measures ANOVA (two-tailed) showed that, across trial types, response time was significantly longer on AY trials relative to AX, BX, and BY trials (in both blocks), and on long-delay trials only, response time was longer on AX trials relative to BX trials and BY trials (all $p < .001$, Bonferroni-corrected). Across delay lengths, response time was significantly longer on short-delay trials relative to long-delay trials for BX trials and BY trials ($p = .01$ and $p = .03$, respectively, Bonferroni-corrected). A 2 × 2 repeated measures ANOVA revealed a significant main effect of cue type, $F(1, 23) =$

[19] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

153.67, $p < .001$, $\eta_p^2 = .87$. There was also a significant main effect of delay-length, $F(1, 23) = 5.67$, $p = .03$, $\eta_p^2 = .20$. The cue type × delay length interaction was significant, $F(1, 23) = 7.65$, $p = .01$, $\eta_p^2 = .25$; this interaction is plotted in Figure 5.2 (B). Post hoc pairwise comparisons using repeated measures ANOVA (two-tailed) showed that, across cue types, response time was significantly longer on updating cues relative to non-updating cues (on both short-delay and long-delay trials; both $p < .001$, Bonferroni-corrected). Across delay lengths, response time was significantly longer on long-delay cues relative to short-delay cues, but this effect was restricted to non-updating cues ($p = .003$, Bonferroni-corrected).

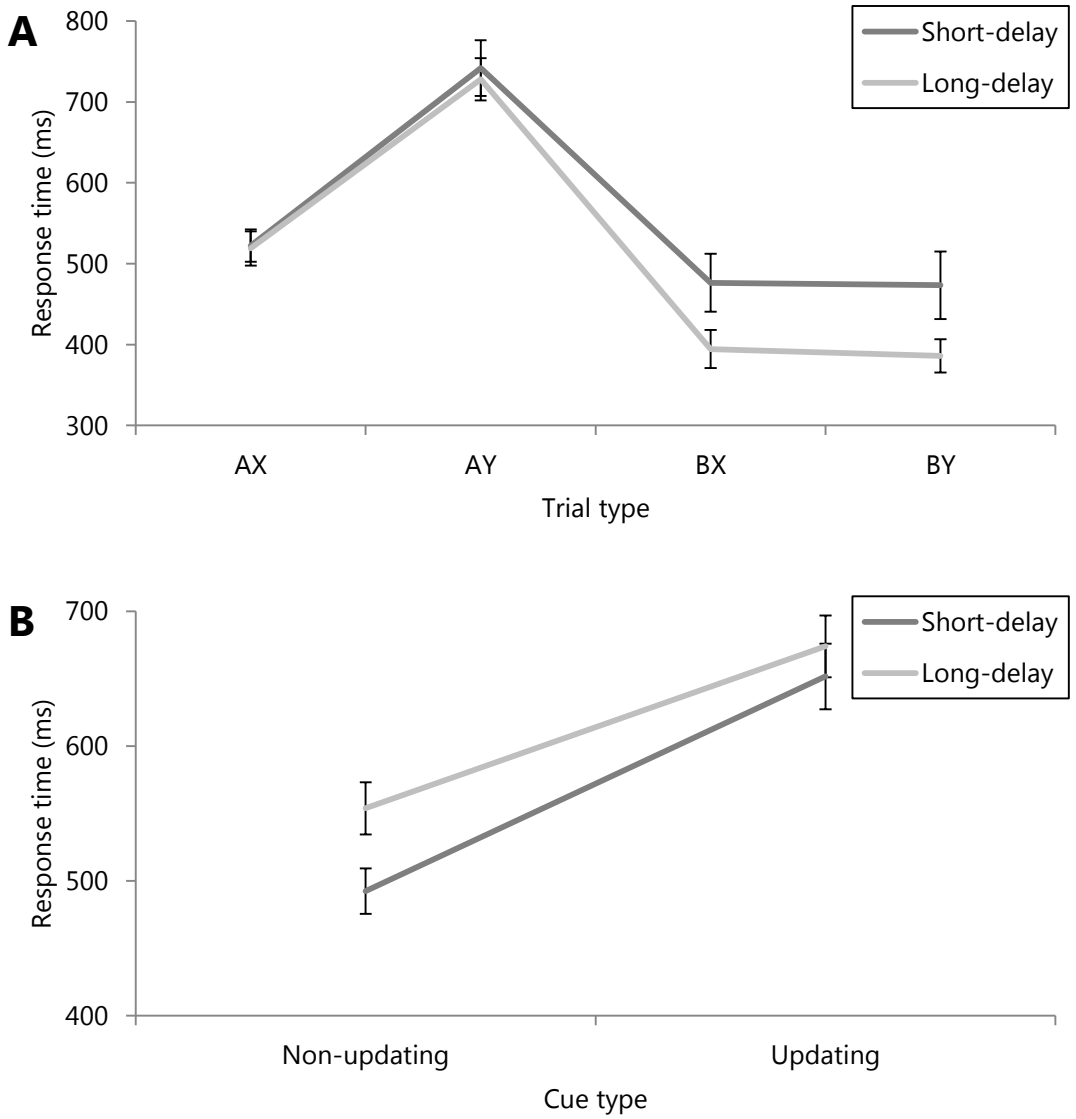


Figure 5.2. Response time (ms) across trial types and delay lengths (A) and across cue types and delay lengths (B). Error bars represent standard error.

***g* correlations.** Pearson's correlations (one-tailed) between Culture Fair error and performance scores are presented in Table 5.4. For a sample size of 24, the critical value for Pearson's *r* (one-tailed) is ±.34 at the $p < .05$ alpha level. Age was not significantly correlated with Culture Fair error, $r(22) = .18$, $p = .40$. Culture Fair error was significantly correlated with: error on updating cues in long-delay, $r(22) = .34$, $p = .05$; error on AX trials in short-delay, $r(22) = .39$, $p = .03$; and response time on BX trials in long-delay, $r(22) = .35$, $p < .05$. All other correlations between performance and Culture Fair error were non-significant ($p > .10$ in all cases). Williams-Hotelling t-

tests (two-tailed) revealed that none of the correlations between performance scores and Culture Fair error differed significantly across cue type, trial type, or delay length ($p > .10$ in all cases).

Table 5.4

Pearson's correlations between Culture Fair error and performance

	Block			
	Short-delay		Long-delay	
Measure	*r*(22)	*p*	*r*(22)	*p*
		Error (%)		
Non-updating	-.06	.40	.001	.50
Updating	.12	.23	.34	.05
AX	.39	.03	.18	.20
AY	.12	.29	.23	.15
BX	.02	.47	.04	.43
BY	.05	.41	-.04	.43
		Response time (ms)		
Non-updating	.05	.40	.06	.38
Updating	.07	.37	.20	.17
AX	.07	.37	-.05	.41
AY	.23	.14	.24	.13
BX	.22	.15	.35	< .05
BY	.19	.19	.24	.13

Culture Fair z-score charts. Participants were assigned to z-score bins (width of .50 SD) based on their Culture Fair raw error scores in order to clarify the relationship between performance scores and Culture Fair performance. Figures 5.3 to 5.6 display performance scores for each measure across Culture Fair z-score bins. The data indicate considerable heterogeneity in performance, and limited evidence for a trend towards better performance in participants at the higher end of the *g* distribution. Nevertheless, consistent with the correlational findings, the charts show trends for reduced error on AX trials in short-delay (Figure 5.3), reduced error on updating cues in long-delay (Figure 5.4), and longer response time on BX trials in long-delay (Figure 5.6) as a

function of Culture Fair score. Participants that scored ≥ 1 SD above the sample mean on the Culture Fair produced 1% fewer AX trials errors (in short-delay) [20], 10% fewer updating cue errors (in long-delay), and responded 96 ms faster to BX trials (in long-delay), than participants that scored ≥ 1SD below the sample mean.

[20] It is acknowledged that a reduction of 1% error is not a noticeable reduction in error in people at the higher end of the *g* distribution relative to people at the lower end of the *g* distribution. However, the small bin sizes, particularly for the bin *>1.01*, decreases the reliability of the data.

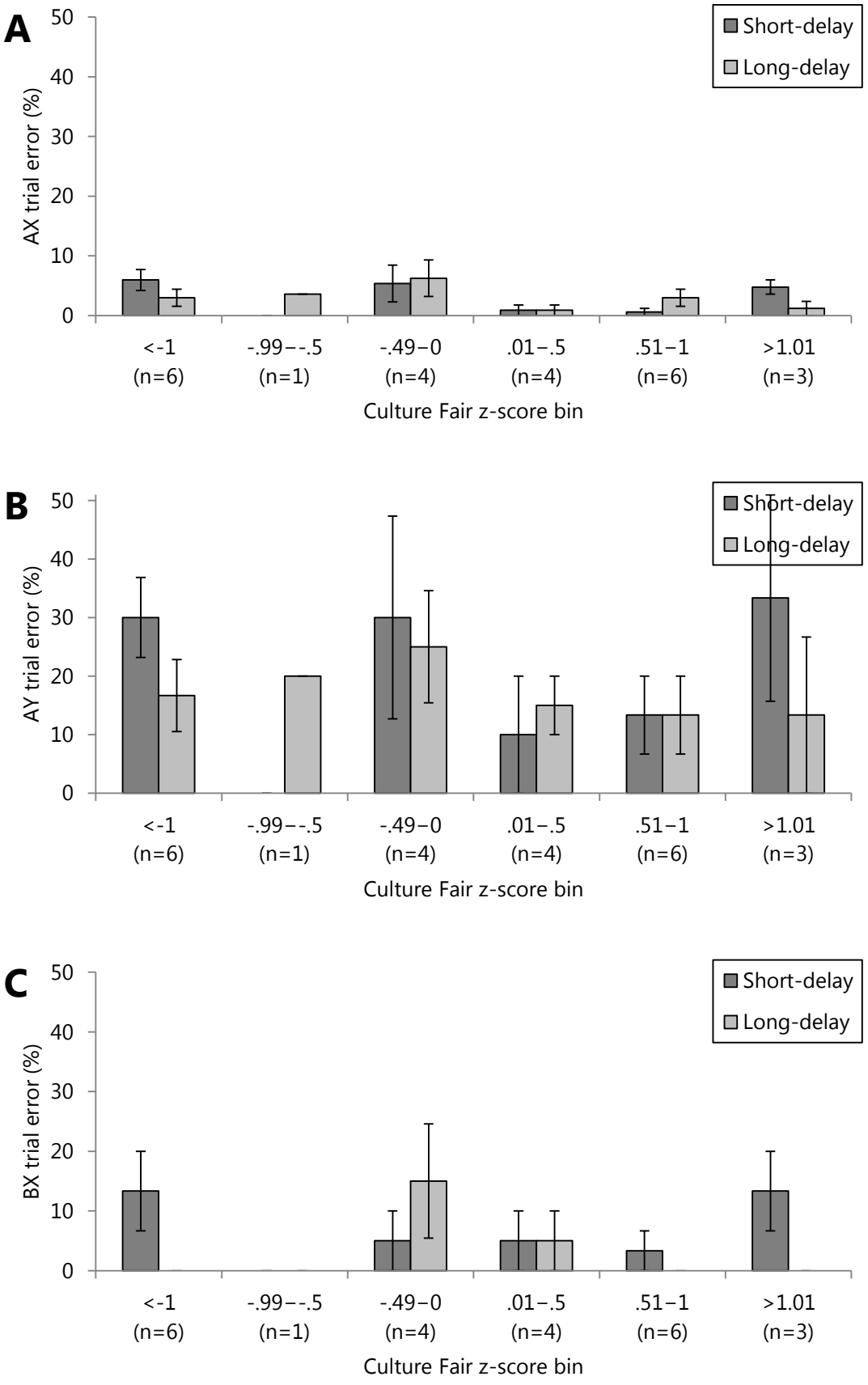


Figure 5.3. Mean error (%) across Culture Fair z-score bins for AX trials (A), AY trials (B), and BX trials (C). Error bars represent standard error.

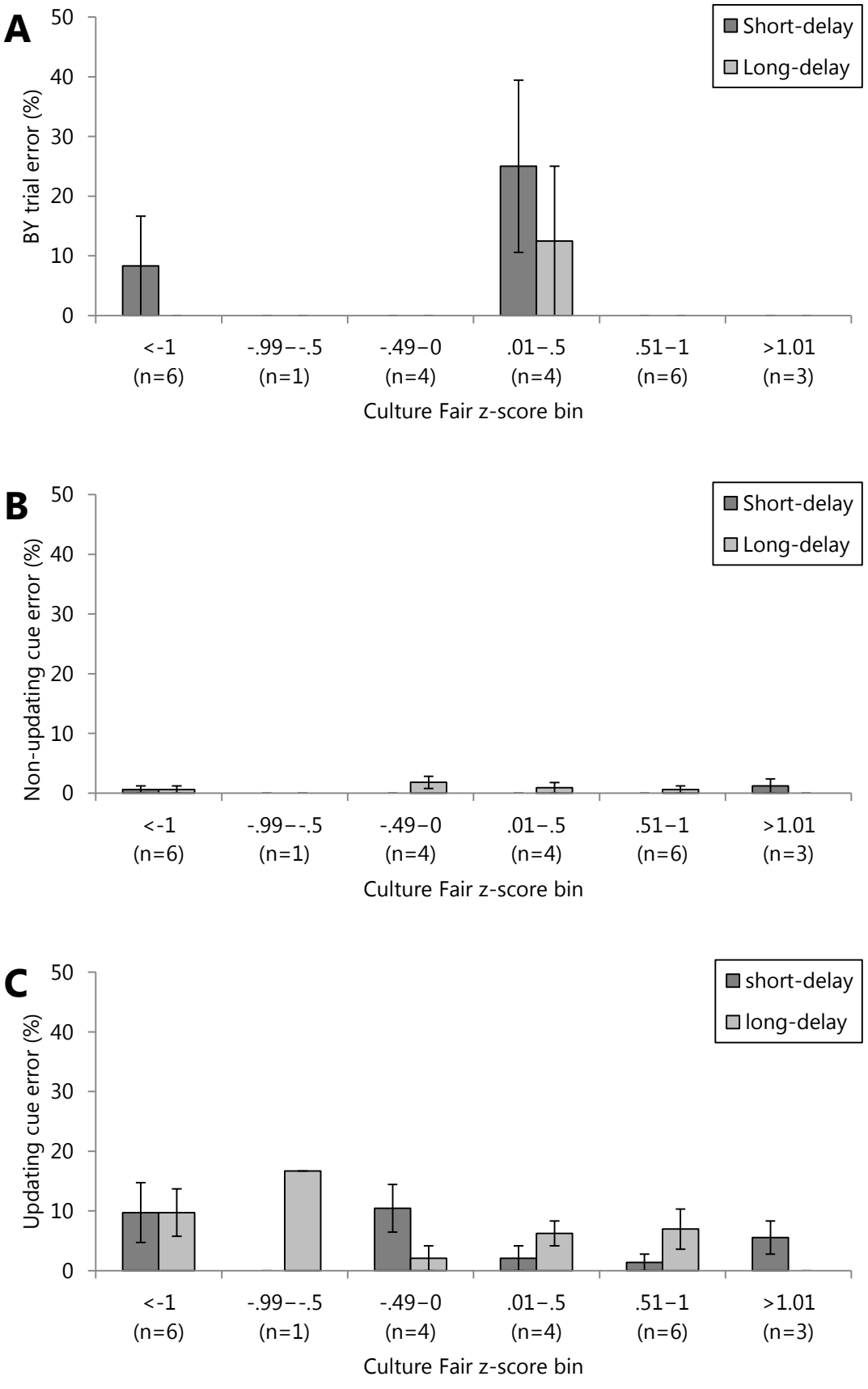


Figure 5.4. Mean error (%) across Culture Fair z-score bins for BY trials (A), non-updating cues (B), and updating cues (C). Error bars represent standard error.

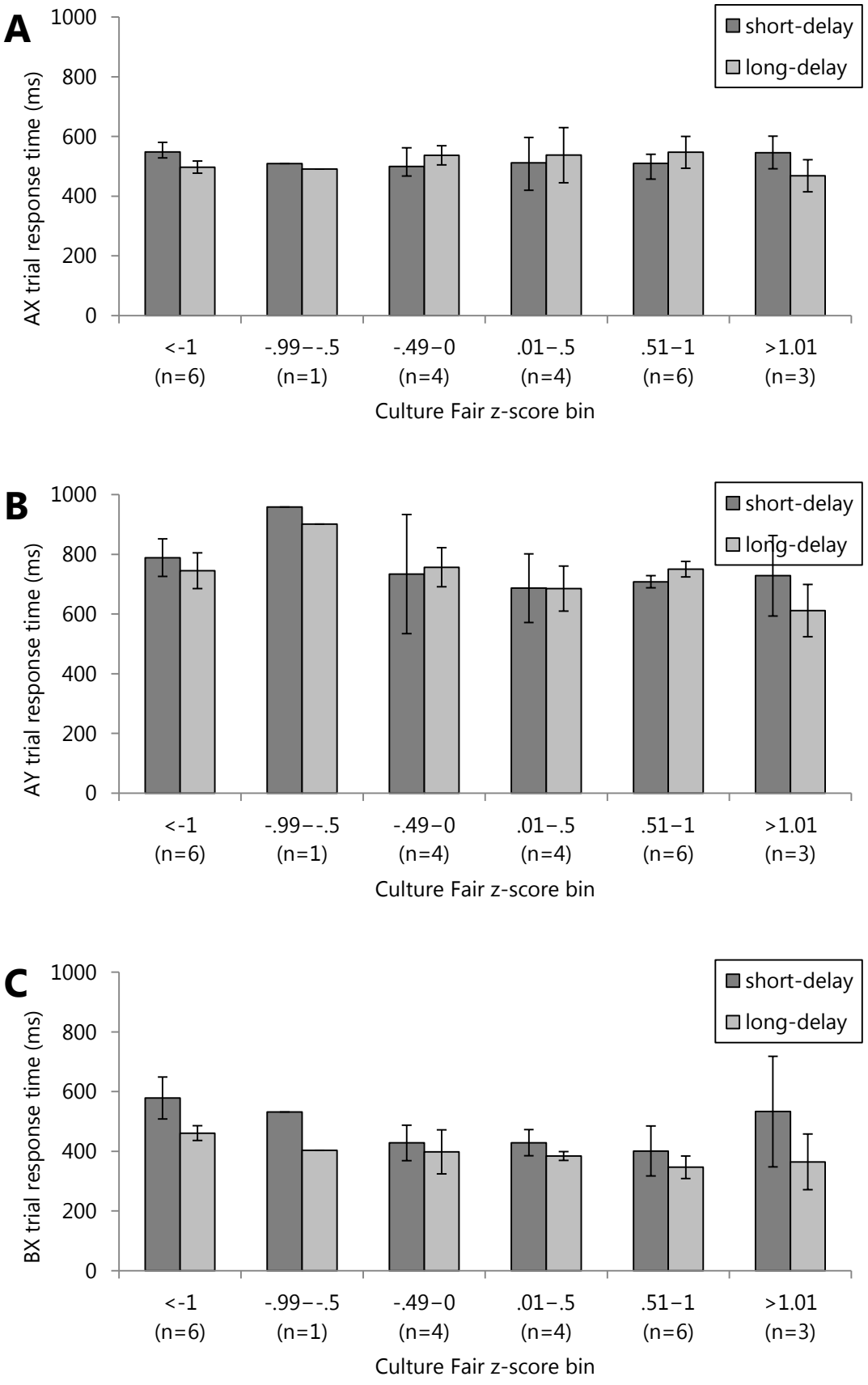


Figure 5.5. Mean response time (ms) across Culture Fair z-score bins for AX trials (A), AY trials (B), and BX trials (C). Error bars represent standard error.

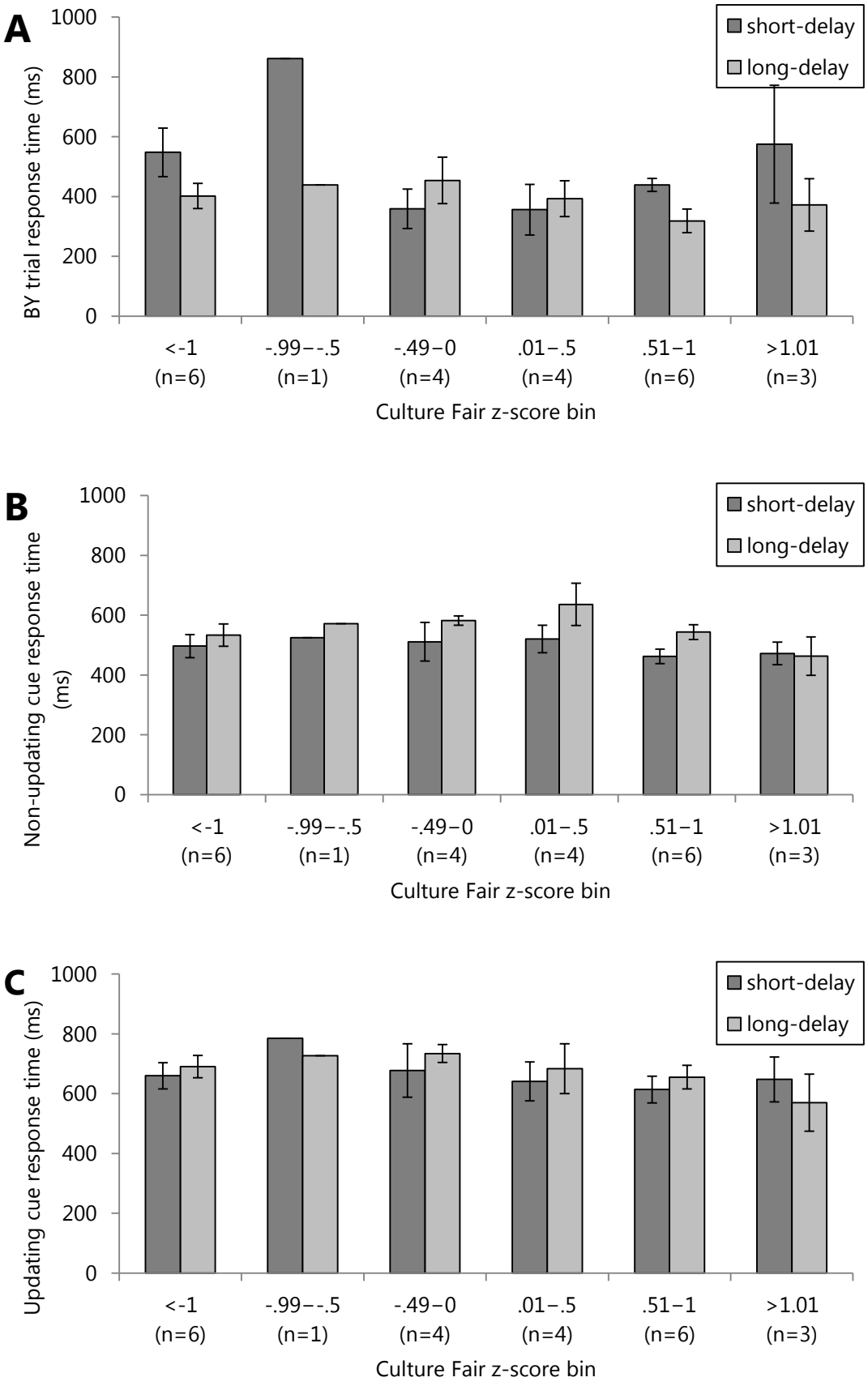


Figure 5.6. Mean response time (ms) across Culture Fair z-score bins for BY trials (A), non-updating cues (B), and updating cues (C). Error bars represent standard error.

Reconceptualisation of task requirements. Based on the number of rules comprised in the selected rule format, which was chosen from the rule sheet, participants were classed as either non-reconceptualisers (if they selected the two-rule format), efficient reconceptualisers (if they selected the one-rule format), or inefficient reconceptualisers (if they selected the three- or the four-rule formats). One-way ANOVA (two-tailed) revealed that Culture Fair error differed significantly across these reconceptualisation groups, $F(2, 21) = 71.67$, $p = .02$, $\eta_p^2 = .33$. Post hoc tests revealed that efficient reconceptualisers ($n = 10$, $M = 6.90$, $SE = 1.18$) produced significantly fewer Culture Fair errors than both non-reconceptualisers ($n = 7$, $M = 11.86$, $SD = 1.86$) and inefficient reconceptualisers ($n = 7$, $M = 11.86$, $SD = 1.86$; both $p = .04$, Bonferroni-corrected), who produced the same number of Culture Fair errors. However, when grouping participants based on whether they reconceptualised task requirements (selected the one-, three, or four-rule formats; $n = 17$, $M = 8.94$, $SD = 4.15$) or not (selected the two-rule format; $n = 7$, $M = 11.86$, $SD = 4.45$), Culture Fair error was statistically equivalent, $t(22) = -1.54$, $p = .14$, $d = .68$.

Performance scores (presented in Table 5.5) were compared across non-reconceptualisers, efficient reconceptualisers, and inefficient reconceptualisers using ANCOVAs (two-tailed) in order to control for the significant difference in Culture Fair scores between groups. Error scores did not differ between groups (all $p > .10$), however, response time for updating cues in long-delay did ($p = .02$). Post hoc pairwise comparisons (two-tailed) revealed that response time for updating cues in long-delay was significantly faster in efficient reconceptualisers relative to both non-reconceptualisers ($p = .05$, Bonferroni-corrected) and inefficient reconceptualisers ($p = .03$, Bonferroni-corrected).

Table 5.5

Performance scores for reconceptualisation groups

	Efficient reconceptualisers		Non-reconceptualisers		Inefficient reconceptualisers	
Measure	*M*	*SE*	*M*	*SE*	*M*	*SE*
			Error (%)			
			Short-delay			
Non-updating	.33	.38	.02	.42	.53	.42
Updating	3.51	3.00	7.02	3.32	7.02	3.32
AX	2.63	1.38	5.01	1.52	2.46	1.52
AY	18.86	8.13	33.68	8.99	13.68	8.99
BX	5.44	4.39	8.97	4.86	8.97	4.86
BY	5.19	6.48	7.01	7.17	7.01	7.17
			Long-delay			
Non-updating	.70	.56	1.03	.63	.52	.62
Updating	6.92	2.67	4.58	2.96	6.96	2.96
AX	2.41	1.40	3.89	1.55	2.87	1.55
AY	19.06	5.40	20.67	5.98	9.24	5.98
BX	-1.27	3.38	9.48	3.74	3.77	3.74
BY	-1.41	3.65	8.15	4.04	1.01	4.04
			Response time (ms)			
			Short-delay			
Non-updating	463	30	490	33	537	33
Updating	597	42	663	47	720	47
AX	489	37	550	40	542	40
AY	724	62	780	69	729	69
BX	464	66	479	73	491	73
BY	532	74	390	82	473	82
			Long-delay			
Non-updating	493	32	580	36	596	36
Updating	587	35	732	39	741	39
AX	466	37	562	41	551	41
AY	674	45	752	50	781	50
BX	374	41	424	45	394	45
BY	363	36	429	40	375	40

Note. Means are adjusted to control for Culture Fair error.

Real-time task execution demand versus task conceptualisation complexity.

Table 5.6 presents Spearman's rank order correlations (one-tailed) between performance scores and the number of rules in the selected rule format (which was assumed to reflect the number of chunks that task requirements were represented in the participants' minds). The number of rules in the selected rule format was significantly correlated with response time on updating cues (in both blocks) and non-updating cues (in long-delay only), but was not significantly correlated with error for any measure. Thus, holding fewer rules in mind is associated with faster responses to cues.

Table 5.6

Spearman's correlations between the number of rules in the selected rule format and performance

	Error (%)		Response time (ms)	
Measure	$r_s(22)$	*p*	$r_s(22)$	*p*
		Short-delay		
Non-updating	.04	.44	.25	.12
Updating	.13	.28	.35	< .05
AX	.25	.12	.19	.19
AY	-.02	.47	.07	.36
BX	.05	.42	.22	.16
BY	.06	.39	-.02	.46
		Long-delay		
Non-updating	-.04	.43	.47	.01
Updating	.27	.10	.57	.002
AX	.15	.25	.16	.22
AY	-.08	.36	.33	.06
BX	.22	.15	.25	.12
BY	.05	.41	.18	.21

The number of rules in the selected rule format was significantly correlated with Culture Fair error, $r_s(22) = .58$, $p = .003$. A series of partial correlations were performed between Culture Fair error and performance scores (when controlling for the number of rules in the selected rule format) and between Culture Fair error and the

number of rules selected (when controlling for the each performance measure separately). None of these correlations differed in strength from their bivariate equivalent (all $p > .10$).

Additional findings: Order effects

Performance. Performance scores are compared across order groups (short–long vs. long–short) in Table 5.7. As was found for the colour shape match task, there was a trend for worse performance in the block that was completed first. Independent-samples t-tests (two-tailed) showed that short-delay error was significantly greater in the short–long group for updating cues and BX trials ($p = .001$ and $p < .001$, respectively, each significant using the Bonferroni-corrected alpha level of $p < .05 / 6 = .004$). Response time did not differ across order groups for any measure (using the Bonferroni-corrected alpha level, $p = .004$).

Table 5.7

Performance scores across order groups

Measure	Order group: Short–long *M*	Short–long *SD*	Long–short *M*	Long–short *SD*	Independent t-test *df*	*t*	*p*	Cohen's *d*
Culture Fair	8.83	4.26	10.75	4.41	22	-1.08	.29	-.44
			Error (%)					
			Short-delay					
Non-updating	.60	1.39	.00	.00	11	1.48	.17	-
Updating	11.11	8.21	.00	.00	11	4.69	.001	-
AX	4.17	3.98	2.38	4.12	22	1.08	.29	.44
AY	30.00	24.86	13.33	17.75	22	1.89	.07	.78
BX	15.00	12.43	.00	.00	11	4.18	< .001	-
BY	8.33	19.46	4.17	14.43	22	.60	.56	.25
			Long-delay					
Non-updating	.60	1.39	.89	1.61	22	-.48	.63	-.19
Updating	2.78	4.10	9.72	8.58	22	-2.53	.02	-1.1
AX [a]	3.27	4.43	2.68	3.09	15.77	.38	.71	.16
AY	18.33	15.86	15.00	15.08	22	.53	.60	.22
BX	3.33	7.78	3.33	11.55	22	.00	1.00	.00
BY	4.17	14.43	.00	.00	11	1.00	.33	-
			Response time (ms)					
			Short-delay					
Non-updating	524	91	461	63	11	1.98	.06	.82
Updating	535	100	573	89	11	1.31	.20	-.40
AX	526	92	519	109	22	.18	.86	.07
AY	742	189	741	152	22	.02	.99	.01
BX	506	215	447	128	11	.81	.43	.34
BY	470	222	476	196	22	-.07	.95	-.03
			Long-delay					
Non-updating	683	125	620	110	11	-.97	.34	.54
Updating	660	688	688	95	11	-.59	.56	.28
AX	493	81	545	127	22	-1.23	.23	-.5
AY	711	157	745	96	22	-.64	.53	-.27
BX	374	115	414	117	11	-.85	.41	-.35
BY	393	97	379	108	22	.35	.73	.14

Note. $n = 12$ in each order group. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

***g* correlations.** Pearson's correlations between performance and Culture Fair error are compared across order groups in Table 5.8. In the short–long order group, Culture Fair error was significantly correlated with performance on updating cues (specifically, error in short-delay and response time in long-delay). In the long–short order group, Culture Fair error was significantly correlated with error on AX trials in both blocks and AY trials in long-delay, and with response time on BY trials in short-delay. Significant correlations ranged between $r = .53$ and $r = .75$. Fisher's z-test (two-tailed) revealed that the correlation between Culture Fair error and BY trial response time in short-delay was stronger in the long–short order group ($p = .03$).

Table 5.8

Pearson's correlations between Culture Fair error and performance across order groups

	Short–long		Long–short		Fisher's z-test	
Measure	*r*(10)	*p*	*r*(10)	*p*	*z*(22)	*p*
			Error (%)			
			Short-delay			
Non-updating	.02	.48	-	-	-	-
Updating	.67	.01	-	-	-	-
AX	.16	.30	.75	.003	1.72	.09
AY	.12	.36	.37	.12	.57	.57
BX	.33	.15	-	-	-	-
BY	.24	.23	-.13	.35	.80	.43
			Long-delay			
Non-updating	.13	.35	-.15	.32	-.60	.55
Updating	.16	.31	.35	.13	.43	.67
AX	-.02	.48	.53	.04	1.29	.20
AY	.14	.33	.66	.01	1.38	.17
BX	.07	.41	.02	.48	.11	.92
BY	.01	.49	-	-	-	-
			Response time (ms)			
			Short-delay			
Non-updating	.33	.15	-.08	.41	.90	.37
Updating	.23	.24	.05	.44	.39	.35
AX	.26	.21	-.06	.42	.69	.49
AY	.27	.20	.20	.27	.16	.88
BX	.26	.21	.31	.16	.12	.91
BY	-.24	.23	.66	.01	2.20	.03
			Long-delay			
Non-updating	.28	.19	-.26	.21	1.18	.24
Updating	.53	.04	-.27	.20	1.84	.07
AX	.28	.19	-.33	.15	-1.34	.18
AY	.45	.07	-.13	.34	1.31	.19
BX	.46	.07	.20	.27	.63	.53
BY	.15	.32	.35	.13	.46	.65

Note. $n = 12$ in each order group.

Discussion

Experiment 4 investigated the recruitment of *g* in performance on a task that placed varying levels of demand on maintaining information in working memory, updating information in working memory, and inhibition in the context of a relatively efficient task model. In accordance with previous experiments, trials that required the inhibition of a prepotent response tendency were particularly difficult for participants. Interestingly, this was restricted to trials in which a *prepared* prepotent response was inappropriate (AY trials, that is, a target cue followed by a non-target probe); error was significantly greater, and response time was significantly slower, on AY trials compared to all other trials. Trials that required the inhibition of a prepotent response to X probes (BX trials, that is, a non-target cue followed by a target probe) were actually easier than target AX trials which placed no demand on inhibition; response time was significantly slower on AX trials compared to BX trials (and neutral BY trials).

These findings may be explained, in part, by differences between A and B trials in the time within a trial that a decision about the appropriate response can be made. Because B cues always indicate a non-target response, a decision regarding the appropriate response can be made early-on in the trial (at the presentation of the cue). Provided that the B cue, or the prepared non-target response, is maintained in working memory across the delay, B trials will be relatively easy regardless of whether the probe is prepotent (X) or not (Y). Conversely, A cues indicate a target response most, but not all, of the time, so a decision regarding the appropriate response cannot be made until the presentation of the probe. On these grounds, responses to AX trials may take longer relative to B trials because the correct response is not indicated until the time of the probe; responses to AY trials may take longer still (and be more error-prone) due to the added requirement to inhibit the prepotent prepared response indicated by the A cue.

Trials that contained a cue that was different to the cue that was presented in the previous trial (e.g., B following A) were also difficult for participants (relative to trials in which the cue was the same as in the previous trial). Error was significantly greater,

and response time was significantly longer, in updating cues relative to non-updating cues. These "updating" cues were classified as requiring the updating of information in working memory for the purpose of this experiment because they appear to involve changing the information to be maintained across a delay. However, it remains unclear whether these cues actually require updating processes akin to those tapped by classical updating tasks (e.g., *n*-back [Gevins & Cutillo, 1993], keep track [Yntema, 1963], and letter-memory [Morris & Jones, 1990] tasks) because the previous cue may decay from working memory at the time of the presentation of the related probe, rather than on the presentation of the current cue. Nonetheless, the finding that updating trials are more difficult than non-updating trials does imply that some kind of extra processing or resource was necessary for them. Further research would need to explore this in more detail to ascertain whether this added processing is akin, or related, to traditional conceptualisations of updating.

Performance differences also emerged between trials that differed in the length of time that the cue was maintained in working memory. For responses to cues (specifically, non-updating cues), response time was longer in the block that required maintenance of the cue across a long (5,000 ms), compared to a short (1,000 ms), delay. One explanation for this could be the shorter inter-trial-interval in the long-delay block, which was incorporated into the design to control for total trial duration; participants have less time in the long-delay block (1,000 ms compared to 5,000 ms) to prepare for making a response to the following cue. That this finding was limited to non-updating cues again suggests that relatively more processing is required to respond to updating cues, even though the same non-target response is required for all cues; this extra processing may increase response time to updating cues regardless of the amount of available time to prepare for the response to the cue. For responses to trials (specifically, BX and BY trials), response time was longer in the short-delay, relative to the long-delay, block. This is inconsistent with the suggestion that demand for working memory maintenance, or context processing (MacDonald et al., 2005), is greater in the

long-delay condition. However, that accuracy on B trials was better, relative to AY trials, supports the notion that context processing is relatively intact in healthy people (MacDonald et al., 2005).

Order effects were also observed. Overall, there was a trend for greater error on trials in short–long participants (who completed the short-delay block first) compared to long–short participants (who completed the long-delay block first), and this finding was not due to group differences in *g* because Culture Fair error was equivalent across order groups. Perhaps this finding can be explained on the basis of differences in difficulty between the two blocks, suggested by the longer response time to some trials (non-updating cues) in the short-delay block. Performance may have been better for long–short participants because their initial experience with the task was with the "easier" version, and so these participants were well-practiced on the task by the time the more "difficult" version was performed. Practice effects were also suggested by the order group comparisons. Short–long participants were less accurate at responding to updating cues and BX trials in the short-delay block relative to long–short participants.

Consistent with the theory that *g* is recruited in tasks that are performed under a complex task conceptualisation, very few task measures were significantly correlated with *g* in Experiment 4 (i.e., when task instructions were given as two distinct rules, imposing a relatively efficient task model). Significant *g* correlations were restricted to accuracy for updating cues in long-delay, accuracy for AX trials in short-delay, and response time for BX trials in long-delay. Moreover, and in line with predictions, *g* correlations did not differ between performance on (a) trials that placed demand on response inhibition and trials that did not, (b) cues that placed demand on "updating" and cues that did not, and (c) trials that placed high demand on maintaining information in working memory and trials that placed low demand on maintenance. These findings support the view that the recruitment of *g* in task performance is unaffected by manipulations of real-time task complexity in the context of an efficient task conceptualisation.

Experiment 4 also sought to investigate reconceptualisation of the task model. Once the task was completed, participants were prompted to select a rule format, from a sheet, that best reflected their own task conceptualisation. The rule format options that participants could choose between differed only in the number of distinct chunks of information (one through four rules) that they comprised; the number of operative task requirements in each format did not differ from the original two-rule task description. Although Culture Fair scores were similar across reconceptualisers (participants that chose one, three, or four rules) and non-reconceptualisers (participants that chose two rules), *efficient* reconceptualisers (participants that chose fewer than two rules) had significantly higher *g* than *inefficient* reconceptualisers and non-reconceptualisers collectively (participants that chose two rules or more). Furthermore, a significant correlation was observed between the number of rules in the chosen rule format and *g*. These findings provide direct evidence for an association between task conceptualisation complexity and Spearman's *g*; people that fall higher on the *g* distribution are more likely to reconceptualise task-relevant information in an efficient way (i.e., by reducing the number of distinct task rules).

The findings also showed that performance on the task was easier under a more efficient task conceptualisation; in the current data set this was restricted purely to response time for cues. Efficient reconceptualisers were significantly faster at responding to updating cues in the long-delay block compared to inefficient and non-reconceptualisers (collectively), and this was independent of differences in Culture Fair error between groups because this was statistically controlled for in these analyses. Furthermore, significant correlations with the number of rules in the rule format chosen were observed for response time for both cue types in long-delay, and also response time for updating cues in short-delay, suggesting that efficient task conceptualisation improves speed of responses.

Another aim of the current experiment was to explore the relative contributions of task-execution versus task-conceptualisation complexity to the recruitment of *g* in task

performance. Partial correlations with *g* were computed for the number of rules chosen (controlling for each performance measure separately), and for performance (controlling for the number of rules in the chosen format), and these partial correlations were compared with their bivariate equivalents. The significant correlation with *g* for the number of rules in the chosen rule format did not change significantly when controlling for any of the performance measures, suggesting that performance did not drive this correlation. The significant correlations with *g* found for some performance measures were also statistically unchanged when controlling for the number of rules in the chosen format, suggesting that the number of rules did not drive any of these correlations. Therefore the type of task complexity (execution vs. conceptualisation) acting to increase task-sensitivity to *g* is unclear, at least under the conditions of this experiment.

In conclusion, although performance was worse in parts of the task that placed relatively more demand on working memory maintenance, working memory updating, and inhibition, correlations with *g* were unaltered by the level of real-time performance demand. These findings extend the conclusions of Experiment 1 and suggest that the recruitment of *g* in task performance is unaffected by manipulations of real-time task complexity in the context of an efficient task conceptualisation. Experiment 4 has slightly advanced understanding of the association between efficient task reconceptualisation and both *g* and task performance by suggesting that efficient task conceptualisation may be restricted to individuals that fall higher on the *g* distribution and may also be associated with better performance on some measures (independently of *g*). However, more data is required to assess the relative contributions of real-time demand versus task conceptualisation complexity to the recruitment of *g* in task performance. In Experiment 5, task instructions are manipulated to form four distinct rules in order to systematically investigate the role of *g* in the context of a more complex task conceptualisation.

CHAPTER 6

Real-time Task Demand, Task Conceptualisation, and Spearman's *g*

Overview

Chapter 6 explores whether the strength of the correlations between DPX task performance and *g* are dependent on the number of distinct task rules presented in task instructions. When task requirements were presented as two rules, in Experiment 4, manipulations of real-time execution demand did not affect task-sensitivity to *g*. Other than the presentation of task requirements as four, instead of two, distinct verbal chunks or rules, the method of Experiment 5 was carried over from Experiment 4, and participants received the same amount of operative task-relevant information. The results showed that, relative to Experiment 4, the recruitment of *g* in task performance was significantly increased, and that the involvement of *g* was significantly higher for conditions that were associated with high, relative to low, real-time execution demand. *g* was also significantly correlated with the number of rules comprised in the rule format selected by participants as representing the number of rules held in mind, and when this measure was statistically controlled, the correlation between inhibition measures and *g* reduced significantly. Collectively, these findings show that task-sensitivity to *g* is increased by types of demand other than prepotent response inhibition, but only in the context of an "inefficient" task model. However, the way in which task-relevant information is conceptualised in mind is the crucial factor in the recruitment of *g*, rather than actual processing demands per se.

Experiment 5

The findings of Experiment 4 showed that the involvement of *g* in DPX task performance was unaffected by manipulations of real-time execution demand—in the form of the inhibition of a prepotent but inappropriate response or a prepared response

(in favour of an alternative response), the maintenance of information across a delay, and the updating of information to be maintained—in the context of two distinct task rules. This was despite greater error on parts of the task that required such demands relative to parts of the task that did not (i.e., trials that did not involve inhibition or updating and trials that involved maintenance across a shorter delay). These findings were consistent with Experiment 1 in which similar patterns of greater error but unchanged correlations with *g* were found on measures involving inhibition relative to non-inhibition measures.

Experiment 5 investigated whether or not increasing the complexity of DPX task instructions, whilst keeping task requirements constant, increases the recruitment of *g* in task performance. The method of splitting each of the two rules of Experiment 1 into two separate rules each in order to form the four rules of Experiment 2 was effective in increasing the relationship between measures involving inhibition and *g*. Thus, an identical technique for splitting two rules into four was used in Experiment 5. This was, predominantly, in order to ascertain whether the finding of increased task-sensitivity to *g*, due to an increase in the number of presented rules, generalises beyond the constraints employed in Experiments 1 through 3. If similar patterns of increased correlations with *g* extend to the demands that are associated with DPX task performance, then it may suggest that the *level*, rather than the *type*, of demand is important. Alternatively, there could be something "special" about response inhibition to *g* as suggested by Experiment 2 (see also Bright, 1998; Dempster, 1991).

Participants in Experiment 5 were also asked to select a rule format that best reflected how the rules were represented in mind. In Experiment 4, the number of distinct task rules comprised in the selected rule formats was related to *g*. The possibility of the observation of strong correlations between real-time demand and *g* in Experiment 5 therefore enabled another focus: the *relative* importance of real-time execution demand versus task modelling to the recruitment of *g* in task performance. Experiment 5 employed a method of comparing (a) observed correlations between task

performance and *g* (e.g., on trials involving inhibition) with partial correlations that statistically controlled for the number of rules in the selected rule format, and (b) the correlation between the number of rules in the selected rule format with partial correlations that statistically controlled for each performance measure in turn. Such comparisons respectively showed whether relationships between aspects of real-time task demand and *g* were driven by the number of rules held in mind, and whether the correlation between the number of rules and *g* was driven by the level of real-time task demand. Performance differences between Experiments 4 and 5 were also analysed in order to help clarify the apparent conflicting findings of the experiments presented thus far and patterns reported in the literature; error was consistent across Experiments 1, 2, and 3 but published studies consistently report increased goal neglect when task instructions are more complex (e.g., Duncan et al., 2008; Dumontheil et al., 2010).

On the basis of the findings of Experiment 2—that the recruitment of *g* increased, in conditions of response inhibition demand, in the context of the presentation of four distinct task rules—it was predicted that the correlation between task performance and Culture Fair error would be significantly stronger on AY and BX trials (inhibition trials) relative to AX and BY trials (non-inhibition trials). However, it was unclear prior to testing whether these effects would also extend to conditions of maintenance and updating demand. Thus, no specific predictions were made concerning whether or not correlations between task performance and Culture Fair error would differ across (a) short-delay and long-delay blocks (in which the cue was maintained for 1,000 ms or 5,000 ms, respectively), or (b) non-updating and updating cues (in which the cue to be maintained was the same as, or different to, the cue in the previously trial, respectively). It was also predicted that (a) correlations between task performance and Culture Fair error would be significantly higher in the four-rule group of Experiment 5 relative to the two-rule group of Experiment 4, and (b) the number of chunks of task rule information comprised in the chosen rule format would be significantly correlated with Culture Fair error.

Method

Participants

Twenty-four adults (17 female) aged between 18 and 44 years (M = 22.46, SD = 6.86) with no history of neurological disorder participated in the study. None of the participants partook in previous experiments. This opportunity sample was a mixture of psychology undergraduate students at Anglia Ruskin University and community volunteers. The number of Culture Fair errors produced by these adults ranged between 2 and 17 (M = 9.17, SD = 4.36). Mean Culture Fair error was statistically equivalent to that produced by the participants in Experiment 4, $t(46) = -.50$, $p = .62$, $d = -.14$. Thus, it is unlikely that any observed differences were due to differences in *g* between rule groups.

Materials and Procedure

Materials (the Culture Fair, the DPX task, and the rule sheet) and procedure were identical to Experiment 4 (see pages 123–126 to 127–128), other than the presentation of four, instead of two, task rules during the verbal administration of task instructions to participants. The first rule was: "when the first pattern appears press red if it is a target." The second rule was: "when the first pattern appears press red if it is a non-target." The third rule was: "when the second pattern appears press green if both patterns were targets." The fourth rule was: "when the second pattern appears press red if either or both patterns were non-targets." See Appendix I for the full research protocol.

Design

Performance measures were identical to those of Experiment 4 (see Table 5.2, page 126). A mixed design was adopted in which performance scores (accuracy and response time) and correlations between performance scores and Culture Fair error were compared across trial types (repeated measures: AX vs. AY vs. BX vs. BY), cue types

(repeated measures: non-updating vs. updating), blocks (repeated measures: short-delay vs. long-delay), order groups (independent samples: short–long [*n* = 12] vs. long–short [*n* = 12]), and rule groups (independent samples: two-rule [Experiment 4] vs. four-rule [Experiment 5]).

Results

Main findings

Performance. Error (%) and response time (ms) scores for cue types (updating, non-updating) and trial types (inhibition: AY and BX; non-inhibition: AX and BY) across delay lengths (short-delay, long-delay) are presented in Table 6.1.

Table 6.1

Mean error and response time across cue type, trial type and delay length

	Block			
	Short-delay		Long-delay	
Measure	*M*	*SD*	*M*	*SD*
		Error (%)		
Non-updating	1.19	2.27	5.95	8.27
Updating	10.76	18.79	18.05	16.79
AX	8.78	13.81	11.16	10.34
AY	30.84	29.47	22.50	23.08
BX	17.50	26.58	18.33	31.71
BY	14.58	27.50	14.58	23.22
		Response time (ms)		
Non-updating	541	96	612	118
Updating	683	109	749	140
AX	592	105	602	131
AY	806	115	799	144
BX	547	160	516	208
BY	560	205	532	250

For error scores, a 4 (trial) × 2 (delay) repeated measures ANOVA (two-tailed) revealed a significant main effect of trial type[21], $F(2.31, 53.06) = 5.09$, $p = .007$, $\eta_p^2 = .18$. There was, however, no main effect of delay length, $F(1, 23) = .09$, $p = .77$, $\eta_p^2 = .004$. The trial type × delay length interaction was also non-significant, $F(3, 69) = 1.55$, $p = .21$, $\eta_p^2 = .06$. Post hoc pairwise comparisons (two-tailed) showed that error was significantly greater on trials requiring the inhibition of a prepared response (AY; $M = 26.67$, $SE = 4.33$) relative to target trials (AX; $M = 9.97$, $SE = 2.07$; $p = .002$, Bonferroni corrected). A 2 (cue) × 2 (delay) repeated measures ANOVA (two-tailed) revealed that there was also a significant main effect of cue type in which updating cues ($M = 14.41$, $SE = 2.54$) were more error-prone than non-updating cues ($M = 3.57$, $SE = .89$), $F(1, 23) = 23.63$, $p < .001$, $\eta_p^2 = .51$. For the cue data, the effect of delay length was also significant, with greater error on cues in the long-delay block ($M = 12.00$, $SE = 2.23$) relative to cues in the short-delay block ($M = 5.98$, $SE = 2.02$), $F(1, 23) = 4.28$, $p = .05$, $\eta_p^2 = .16$. There was, however, no cue type × delay length interaction, $F(1, 23) = .25$, $p = .63$, $\eta_p^2 = 01$.

A similar pattern was observed for the response time scores. A 4 × 2 repeated measures ANOVA (two-tailed) revealed a significant main effect of trial type[21], $F(1.91, 36.37) = 45.88$, $p < .001$, $\eta_p^2 = .71$. However, there was no main effect of delay length, $F(1, 19) = .39$, $p = .54$, $\eta_p^2 = .02$. The trial type × delay length interaction was also non-significant, $F(3, 57) = .52$, $p = .67$, $\eta_p^2 = .03$. Post hoc pairwise comparisons (two-tailed) showed that response time was significantly slower on trials requiring the inhibition of a prepared response (AY; $M = 802.64$, $SE = 26.28$) relative to target trials (AX; $M = 597$, $SE = 23.29$) and to trials associated with the inhibition of a prepotent response to the probe (but also associated with an *appropriate* prepared response; BX; $M = 531.07$, $SE = 37.81$), and BY trials ($M = 545.98$, $SE = 46.34$; all $p < .001$, Bonferroni corrected). A 2 × 2 repeated measures ANOVA (two-tailed) revealed a significant main effect of cue type, $F(1, 23) = 108.39$, $p < .001$, $\eta_p^2 = .83$; response time

[21] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

was longer on updating cues (M = 716.39, SE = 23.20) relative to non-updating cues (M = 576.53, SE = 19.29). There was also a significant effect of delay length, $F(1, 23)$ = 13.75, p = .001, η_p^2 = .37; response time was longer on updating cues (M = 716.39, SE = 23.20) relative to non-updating cues (M = 576.53, SE = 19.29). However, there was no cue type × delay length interaction, $F(1, 23)$ = .07, p = .79, η_p^2 = .003.

***g* correlations.** Pearson's correlations (one-tailed) between Culture Fair error and performance scores are presented in Table 6.2. A number of performance measures were significantly correlated with Culture Fair error, in particular: error on BX trials and BY trials in the short-delay block, and error on non-updating cues, updating cues, and both types of inhibition trial (AY and BX) in long-delay block; significant correlations ranged between r = .40 and r = .66. Marginal correlations with Culture Fair error were observed for error on: updating cues and AY trials in short-delay, each $r(22)$ = .31, p = .07; AX trials in long-delay, $r(22)$ = .31, p = .07; and BY trials in long-delay, $r(22)$ = .30, p = .08. For response time, Culture Fair error was significantly correlated with updating cues in long-delay, AX trials in long-delay, BX trials in short-delay, and BY trials in both blocks; significant correlations ranged between r = .34 and r = .46. Marginal correlations were observed for: non-updating cues in long-delay, $r(22)$ = .29, p = .09; and BX trials in long-delay, $r(22)$ = .33, p = .07. Williams-Hotelling t-tests (two-tailed) revealed that the correlation with Culture Fair error was significantly stronger for error on: (a) updating cues in the long-delay block relative to the short-delay block, $t(21)$ = 2.24, p = .04; (b) updating cues relative to non-updating cues (in short-delay), $t(21)$ = 2.25, p = .04; (c) BX trials relative to AX trials (in short-delay), $t(21)$ = 2.38, p = .02; and (d) BY trials relative to AX trials (in short-delay), $t(21)$ = 2.57, p = .02. However, none of the correlations between Culture Fair error and response time differed as a function of cue type, trial type or delay length (all p > .10).

Table 6.2

Pearson's correlations between Culture Fair error and performance

	Block			
	Short-delay		Long-delay	
Measure	*r*(22)	*p*	*r*(22)	*p*
		Error (%)		
Non-updating	.29	.09	.42	.02
Updating	.31	.07	.40	.03
AX	.08	.35	.31	.07
AY	.31	.07	.62	.001
BX	.56	.002	.66	< .001
BY	.51	.006	.30	.08
		Response time (ms)		
Non-updating	.13	.28	.29	.09
Updating	.27	.10	.36	.04
AX	.12	.30	.34	.05
AY	.28	.11	.24	.13
BX	.46	.01	.33	.07
BY	.41	.03	.37	.04

Culture Fair z-score charts. Performance scores for each measure across Culture Fair z-score bins (width of .50 SD) are presented in Figures 6.1 to 6.4. The data demonstrate a strong trend towards better performance, as displayed by both fewer errors and faster response speed, in participants at the higher end of the *g* distribution. Participants that scored ≥ 1 SD above the sample mean on the Culture Fair produced 19% fewer updating cue errors (in long-delay), 55% fewer AY trial errors (in long-delay), 48-65% fewer BX trial errors (across blocks), and 37.5% fewer BY trial errors (in short-delay), than participants that scored ≥ 1 SD below the sample mean. Participants that scored ≥ 1 SD above the sample mean on the Culture Fair also took 140 ms less time to respond to updating cues (in long-delay), 139 ms less time to respond to AX trials (in long-delay), 274–342 ms less time to respond to BX trials (across blocks), and 429–465 ms less time to respond to BY trials (across blocks), than participants that scored ≥ 1 SD below the sample mean.

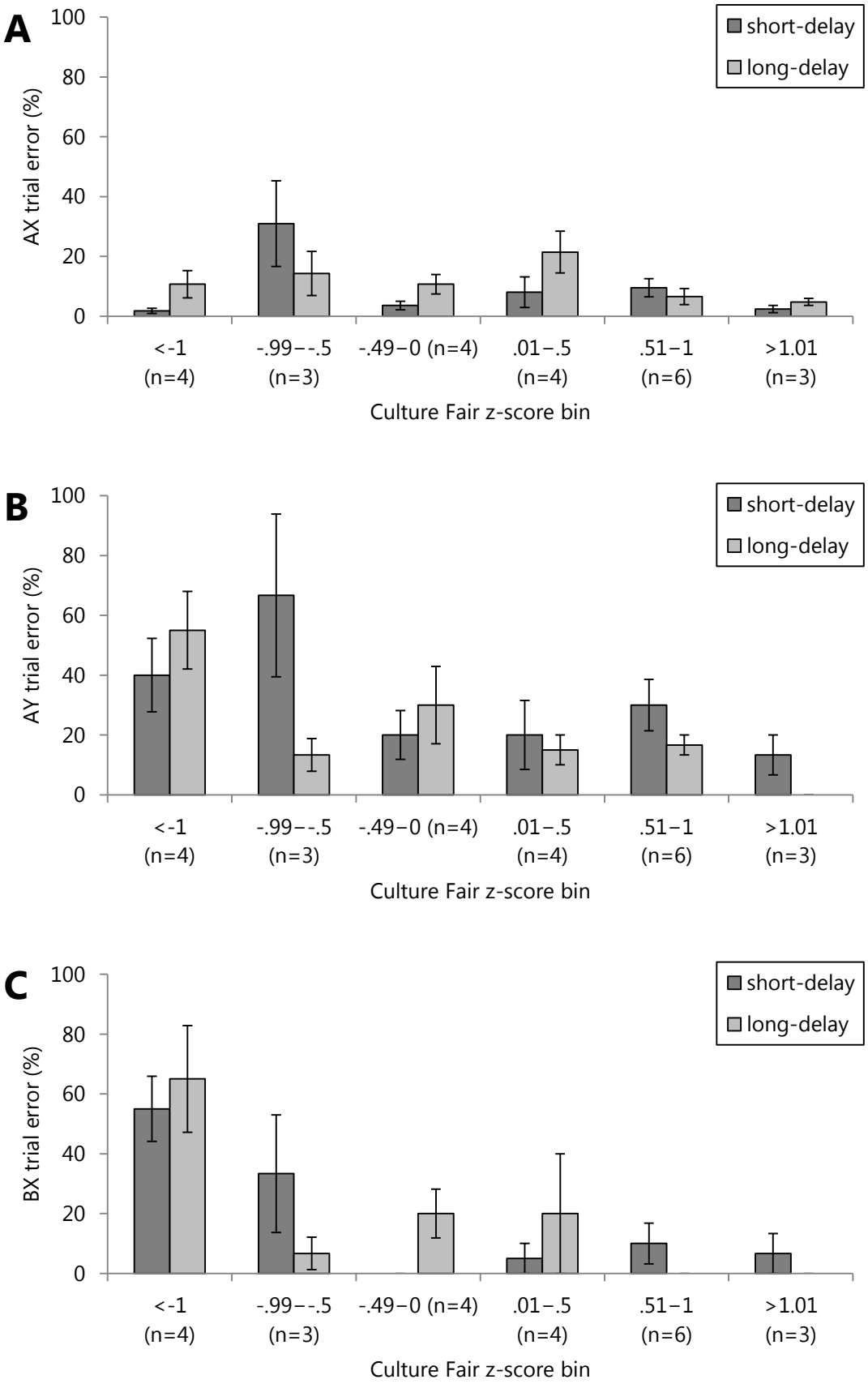


Figure 6.1. Mean error (%) across Culture Fair z-score bins for AX trials (A), AY trials (B), and BX trials (C). Error bars represent standard error.

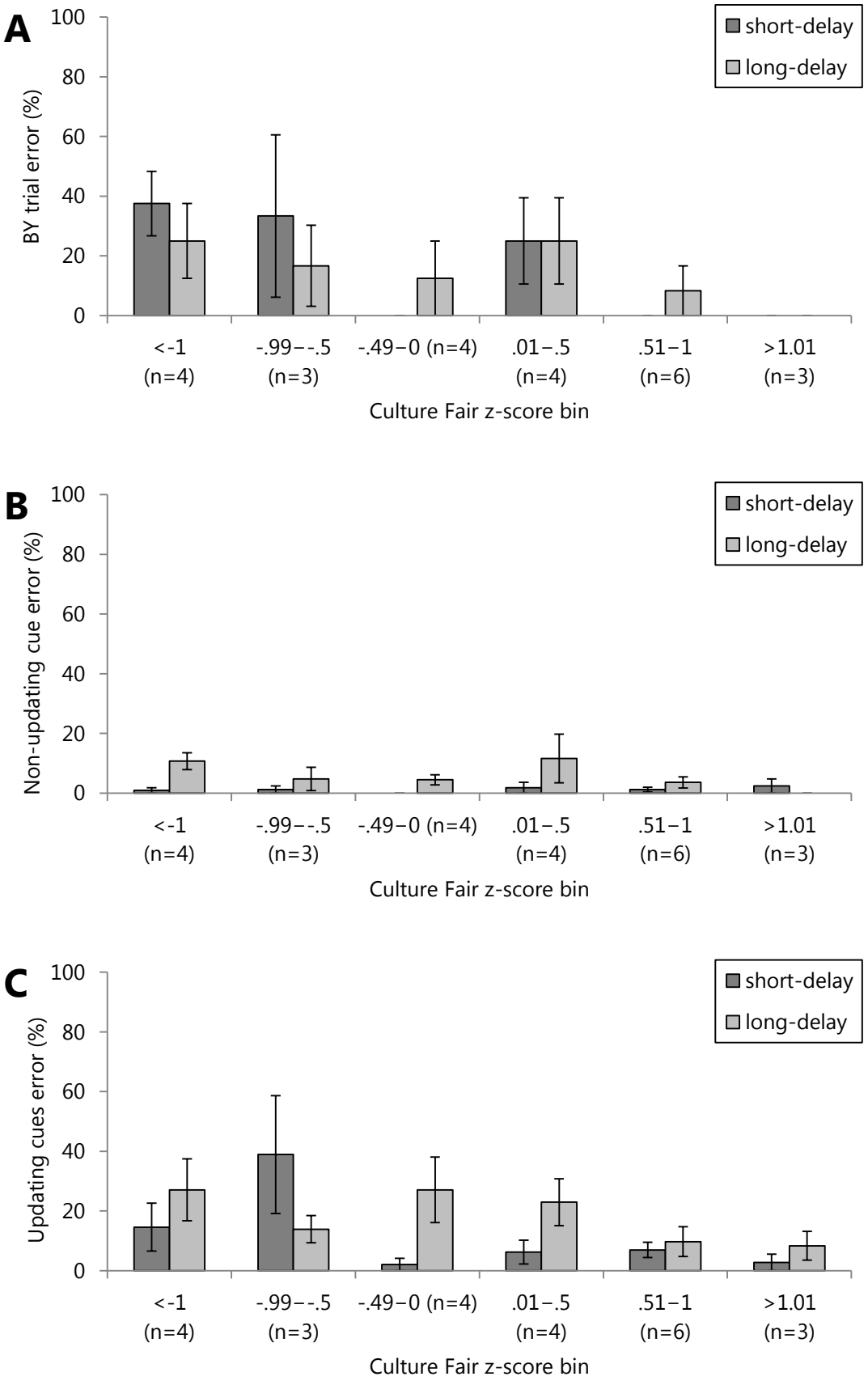


Figure 6.2. Mean error (%) across Culture Fair z-score bins for BY trials (A), non-updating cues (B), and updating cues (C). Error bars represent standard error.

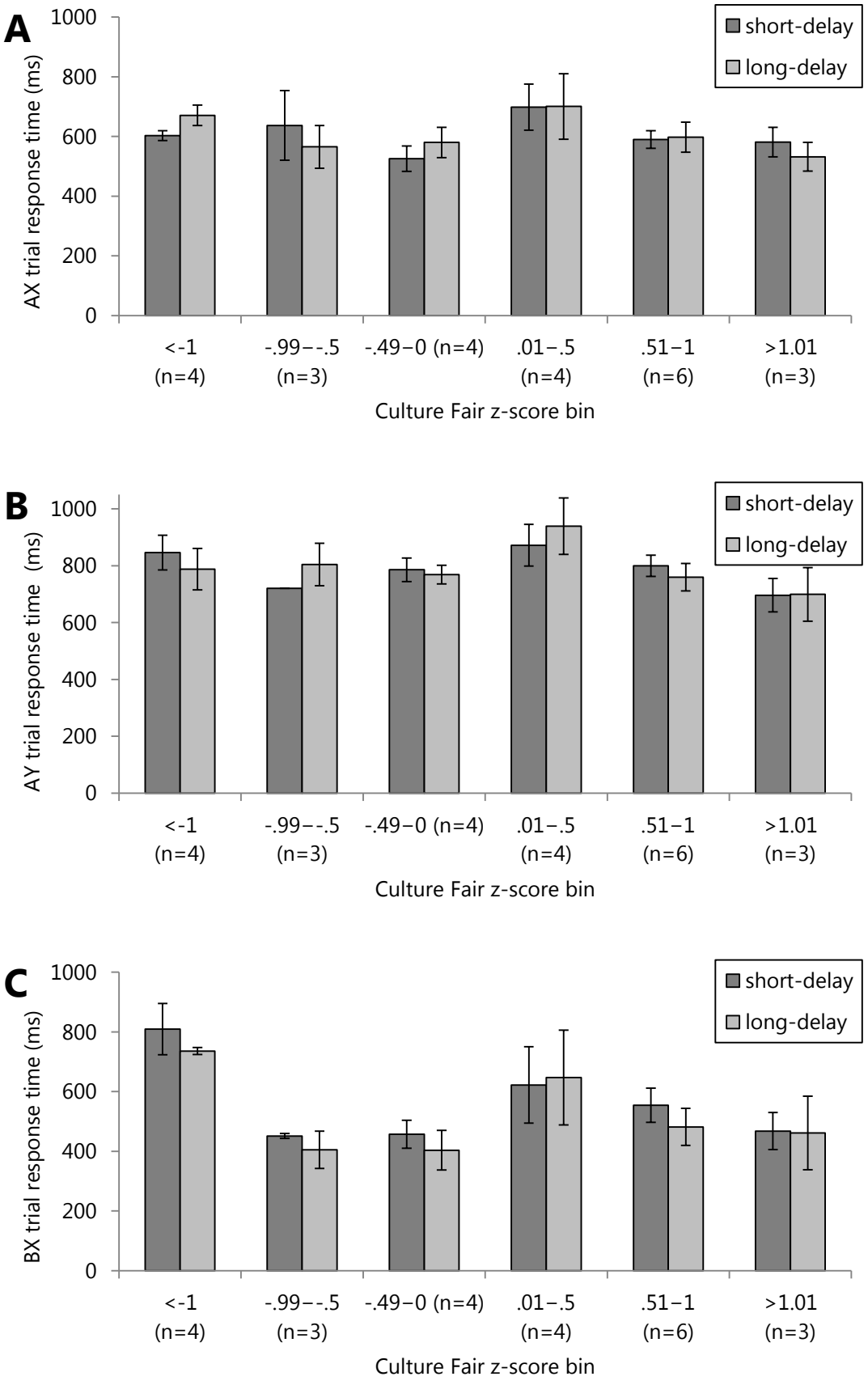


Figure 6.3. Mean response time (ms) across Culture Fair z-score bins for AX trials (A), AY trials (B), and BX trials (C). Error bars represent standard error.

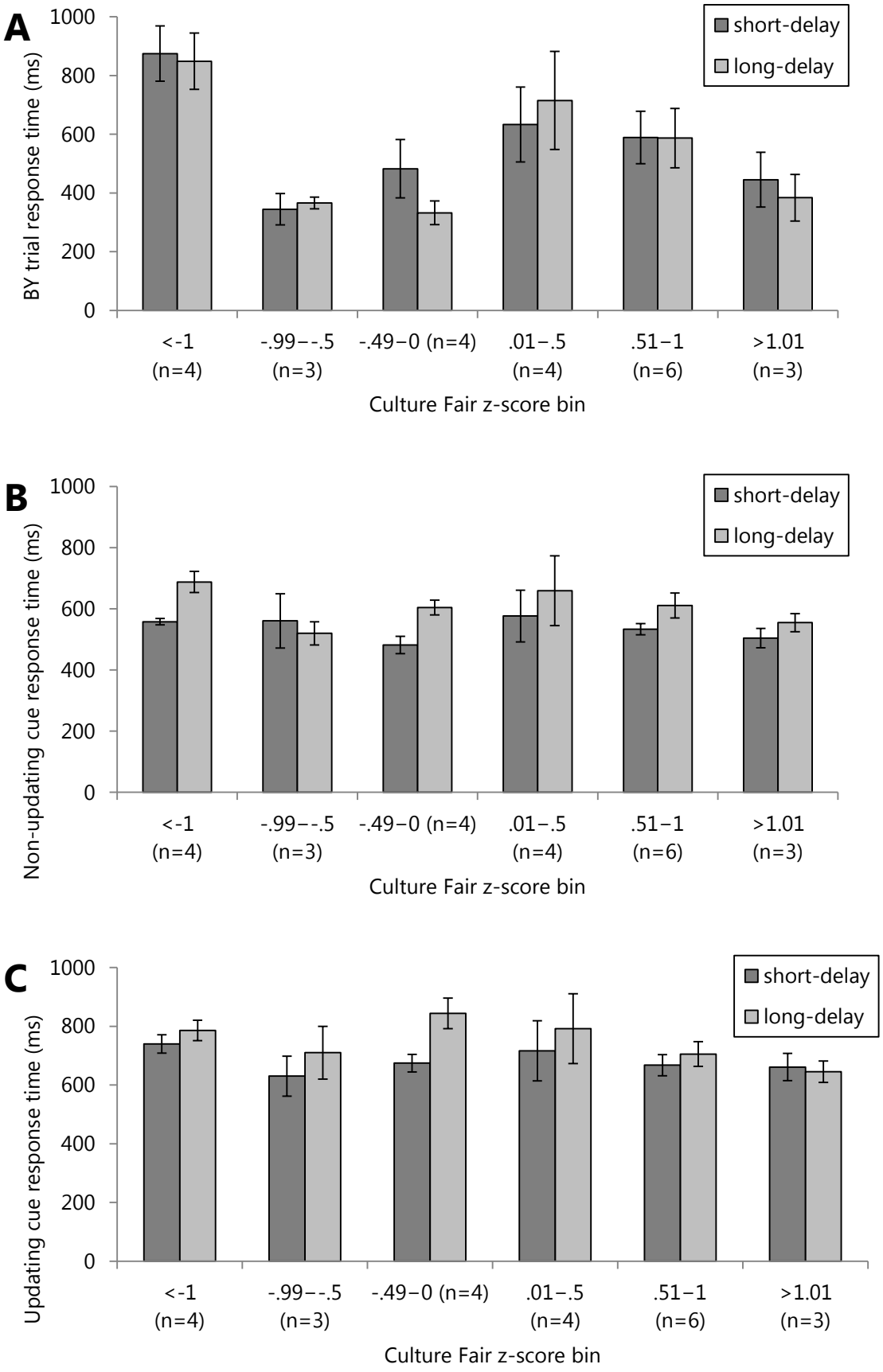


Figure 6.4. Mean response time (ms) across Culture Fair z-score bins for BY trials (A), non-updating cues (B), and updating cues (C). Error bars represent standard error.

Comparisons across rule groups. Table 6.3 compares performance levels (independent samples t-tests, two-tailed) and correlations between Culture Fair error and each performance measure (Fisher's z-tests, two-tailed) across rule groups. Independent samples t-tests showed that, in the four-rule group (Experiment 5) relative to the two-rule group (Experiment 4), error was significantly greater on updating cues, non-updating cues, non-inhibition trials (AX and BY) and inhibition trials (BX), but this was restricted to the long-delay block. Response time was significantly longer on updating cues (in long-delay), inhibition trials (BX in long-delay), and non-inhibition trials (AX in both blocks and BY in long-delay). Fisher's z-tests revealed that the correlation with Culture Fair error was significantly higher in the four-rule group relative to the two-rule group for error on inhibition trials (BX in both blocks). However, none of the correlations between response time and *g* differed across rule groups (all $p > 10$).

Table 6.3

Performance (independent samples t-tests) and correlations between Culture Fair error and performance (Fisher's z-tests) across rule groups

	Independent t-test			Cohen's *d*	Fisher's z-test	
Measure	*df*	*t*	*p*		*z*(46)	*p*
			Error (%)			
			Short-delay			
Non-updating [a]	31.70	-1.76	.09	.54	1.16	.25
Updating	46	-1.25	.22	.39	.65	.52
AX [a]	26.96	-1.87	.07	.62	1.08	.28
AY	46	-1.21	.23	.35	.65	.52
BX [a]	31.34	-1.69	.10	.53	1.99	< .05
BY [a]	38.19	-1.27	.21	.38	1.66	.10
			Long-delay			
Non-updating [a]	24.47	-3.04	.006	1.07	1.45	.15
Updating [a]	31.78	-3.15	.004	.97	.23	.82
AX [a]	28.95	-3.65	.001	1.16	.45	.65
AY	46	-1.03	.31	.30	1.59	.11
BX [a]	27.21	-2.22	.04	.73	2.44	.01
BY [a]	31.57	-2.42	.02	.75	1.13	.26
			Response time (ms)			
			Short-delay			
Non-updating	46	-1.87	.07	.54	.26	.79
Updating	46	-.96	.34	.28	.67	.50
AX	46	-2.60	.01	.69	.16	.87
AY	44	-1.38	.18	.46	.17	.86
BX	45	-1.81	.08	.42	.89	.38
BY	45	-1.74	.09	.43	.79	.43
			Long-delay			
Non-updating	46	-1.89	.07	.55	.77	.44
Updating	46	-2.05	.05	.60	.56	.57
AX	46	-2.69	.01	.71	1.31	.19
AY	46	-1.63	.11	.52	.00	1.00
BX [a]	32.64	-2.34	.03	.75	.07	.94
BY [a]	29.17	-2.88	.007	.83	.47	.64

Note. [a] Degrees of freedom and *p* adjusted (for the independent t-test) using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

Reconceptualisation of task requirements. Participants were classed as either (a) non-reconceptualisers (if they selected the four-rule format from the rule sheet) or reconceptualisers (if they selected the one-, two-, or three-rule formats), and (b) efficient conceptualisers (if they selected the one- or two-rule formats) or inefficient conceptualisers (if they selected the three- or four-rule formats). Independent samples t-test (two-tailed) showed that reconceptualisers (n = 19, M = 7.47, SD = 3.04) produced significantly fewer Culture Fair errors than non-reconceptualisers (n = 5, M = 15.60, SD = 1.52), $t(22) = -5.72$, $p < .001$, $d = -3.57$. Also, efficient conceptualisers (n = 18, M = 7.22, SD = 2.92) produced significantly fewer Culture Fair errors than inefficient conceptualisers (n = 6, M = 15.00, SD = 2.00), $t(22) = -6.02$ $p < .001$, $d = -3.16$.

Performance scores (error and response time) are compared across efficient conceptualisers and inefficient conceptualiser in Table 6.4; all means are adjusted to control for the significant difference in Culture Fair error between groups. ANCOVAs (two-tailed), controlling for Culture Fair error, revealed that efficient conceptualisers made significantly fewer BX trial errors (in short-delay) than inefficient conceptualisers, but that they also responded significantly *slower* on updating cues (in long-delay).

Table 6.4

Performance scores across efficient and inefficient conceptualisers

	Efficient conceptualisers (n = 18)		Inefficient conceptualisers (n = 6)		ANCOVA		
Measure	*M*	*SE*	*M*	*SE*	$F(1, 22)$	*p*	η_p^2
			Error (%)				
			Short-delay				
Non-updating	.71	.64	2.62	1.39	1.19	.29	.05
Updating	6.37	5.04	23.94	10.99	1.62	.22	.07
AX	7.90	4.02	11.42	8.77	.10	.75	.005
AY	30.59	8.21	31.56	17.90	.002	.97	.00
BX	6.35	5.40	50.94	11.76	9.10	.01	.30
BY	7.06	6.53	37.15	14.24	2.83	.11	.12
			Long-delay				
Non-updating	6.93	2.17	3.01	4.74	.43	.52	.02
Updating	21.03	4.41	9.12	9.62	.97	.34	.04
AX	13.26	2.80	4.86	6.10	1.20	.29	.05
AY	20.95	5.30	27.16	11.54	.18	.67	.009
BX	17.89	7.01	19.66	15.29	.01	.93	.00
BY	13.53	6.49	17.74	14.14	.06	.82	.003
			Response time (ms)				
			Short-delay				
Non-updating	536	24	500	75	.18	.67	.01
Updating	697	27	573	86	1.64	.22	.09
AX	607	28	508	89	.96	.34	.05
AY	815	28	754	89	.37	.55	.02
BX	549	42	536	132	.008	.93	.00
BY	570	55	503	172	.12	.73	.007
			Long-delay				
Non-updating	633	30	457	93	2.80	.11	.14
Updating	787	29	465	92	9.59	.007	.36
AX	635	32	417	100	3.74	.07	.18
AY	826	34	645	107	2.24	.15	.12
BX	533	53	418	165	.38	.55	.02
BY	553	66	412	208	.36	.56	.02

Note. Means are adjusted to control for Culture Fair error.

Real-time task execution demand versus task conceptualisation complexity. Spearman's rank order correlations (one-tailed) between performance scores and the number of rules in the selected rule format are presented in Table 6.5. The number of rules in the selected rule format was significantly correlated with error on inhibition trials (BX) in both delay blocks and non-inhibition trials (BY) in short delay, and with response time on inhibition (AY) and non-inhibition (BX) trials in short-delay. This suggests that holding fewer rules in mind is associated with better performance on a number of measures, especially those measures associated with inhibition.

Table 6.5

Spearman's correlations between the number of rules in the selected rule format and performance

	Error (%)		Response time (ms)	
Measure	*r*(22)	*p*	*r*(22)	*p*
		Short-delay		
Non-updating	-.02	.46	.05	.41
Updating	.21	.16	.17	.22
AX	-.14	.26	-.03	.44
AY	.25	.12	.20	.19
BX	.61	.001	.35	< .05
BY	.45	.01	.28	.10
		Long-delay		
Non-updating	.24	.13	.04	.43
Updating	.13	.27	.15	.24
AX	-.01	.48	.14	.26
AY	.52	.004	.11	.30
BX	.65	< .001	.14	.26
BY	.28	.09	.26	.11

The number of rules represented in the participants' minds was significantly correlated with Culture Fair error, $r_s(22) = .76$, $p < .001$. A series of partial correlations were performed between Culture Fair error and performance scores (when controlling for the number of rules in the selected rule format) and between Culture Fair error and

the number of rules in the selected rule format (when controlling for each performance measure separately). None of the partial correlations between the number of rules and Culture Fair error differed in strength from their bivariate equivalent (all $p > .10$). However, the correlations between Culture Fair error and response time on inhibition trials (in short-delay) were significantly weaker when controlling for the number of rules in the participants' conceptualisations for AY trials ($p = .04$) and BX trials ($p = .003$). Although significant effects were restricted to short-delay trials, marginal effects were observed in long-delay (AY, $p = .09$; BX, $p = .08$).

Additional findings: Order effects

Performance. Performance scores are compared across order groups in Table 6.6. Independent samples t-tests (two-tailed) showed that error was significantly greater in the short–long order group for BX trials in short-delay ($p < .001$), and was significantly greater in the long–short order group for updating cues in long-delay ($p = .001$; each significant using the Bonferroni-corrected alpha level of $p < .05 / 6 = .004$). However, using the Bonferroni-corrected alpha level ($p = .004$), response time did not differ across order groups for any measure.

Table 6.6

Performance scores across order groups

	Order group				Independent t-test			Cohen's
	Short–long		Long–short					
Measure	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	*d*
Culture Fair	8.75	4.81	9.58	4.03	22	-.46	.65	-.19
				Error (%)				
				Short-delay				
Non-updating	2.38	2.78	.00	.00	11	2.97	.01	-
Updating [a]	19.44	23.39	2.08	5.18	12.08	2.51	.03	1.22
AX [a]	14.29	17.89	3.27	3.56	11.87	2.09	.06	1.03
AY	38.33	35.63	23.33	20.60	22	1.26	.22	.53
BX	21.67	26.23	13.33	27.41	22	.76	.46	.31
BY	16.67	32.57	12.5	22.61	22	.36	.72	.15
				Long-delay				
Non-updating	.00	.00	7.44	10.27	22	-.88	.39	-
Updating [a]	2.08	5.18	27.08	18.16	16.13	-3.08	.01	-2.14
AX	8.63	8.68	13.69	11.58	22	-1.21	.24	-.50
AY	18.33	23.29	26.67	23.09	22	-.88	.39	-.36
BX [a]	10.00	15.95	26.67	41.19	14.23	-1.31	.21	-.58
BY [a]	8.33	19.46	20.83	25.75	20.48	-1.34	.19	-.55
				Response time (ms)				
				Short-delay				
Non-updating	562	116	520	68	22	1.09	.29	.46
Updating	662	120	705	98	22	-.98	.34	-.29
AX	583	106	601	109	22	-.01	1.00	-.16
AY	823	145	789	78	22	.83	.42	.31
BX	543	157	550	172	22	-.96	.35	-.05
BY	551	208	569	212	22	-1.21	.24	-.09
				Long-delay				
Non-updating	580	107	645	123	22	-1.39	.18	-.57
Updating	701	128	797	140	22	-1.75	.09	-.72
AX	534	68	671	146	22	-2.95	.01	-.13
AY	746	145	852	127	22	-1.18	.25	-.78
BX	461	198	570	212	22	-1.35	.19	-.53
BY [a]	506	210	557	292	18.27	-1.33	.20	-.20

Note. $n = 12$ in order each group. [a] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to homogeneous variance across groups.

***g* correlations.** Pearson's correlations (one-tailed) between performance and Culture Fair error are compared across order groups in Table 6.7. In the short–long order group, Culture Fair error was significantly correlated with error on updating cues (in short-delay), non-updating cues (in long-delay), AY trials (in both blocks), and BX trials (in both blocks), and with response time on AY trials (in both blocks), and BX trials (in long-delay; a marginal correlation was observed on AX trials in long-delay, $p = .053$). In the long–short order group, Culture Fair error was significantly correlated with error on updating cues (in long-delay), AX trials (in long-delay), AY trials (in long-delay), BX trials (in both blocks), and BY trials (in short-delay), and with response time on BY trials (in short-delay; a marginal correlation was observed on BX trials in short-delay, $p = .052$). Fisher's z-test (two tailed) revealed that none of the correlations between performance and Culture Fair error differed significantly across order groups. Marginal effects were observed for AY trials; the correlation between Culture Fair error and AY trial performance was larger in short–long participants (relative to long–short participants) for error in short-delay ($p = .07$) and for response time in long-delay ($p = .06$).

Table 6.7

Pearson's correlations between Culture Fair error and performance across order groups

	Order group					
	Short–long		Long–short		Fisher's z-test	
Measure	*r*(10)	*p*	*r*(10)	*p*	*z*(22)	*p*
			Error (%)			
			Short-delay trials			
Non-updating	-.19	.27	-	-	-	-
Updating	.55	.03	-.03	.47	1.38	.17
AX	.21	.26	-.19	.28	.86	.39
AY	.60	.02	-.16	.31	1.81	.07
BX	.52	.04	.66	.009	.46	.65
BY	.41	.10	.71	.005	.96	.34
			Long-delay trials			
Non-updating	.52	.04	.40	.10	.32	.75
Updating	.27	.20	.56	.03	.76	.45
AX	.08	.40	.51	< .05	1.02	.31
AY	.65	.01	.58	.02	.24	.81
BX	.72	.004	.76	.002	.19	.85
BY	.12	.35	.44	.08	.75	.46
			Response time (ms)			
			Short-delay trials			
Non-updating	.09	.39	.29	.18	.44	.66
Updating	.15	.32	.41	.09	.60	.55
AX	.16	.32	.05	.44	.24	.81
AY	.55	< .05	-.07	.42	1.46	.14
BX	.43	.09	.49	.05	.16	.87
BY	.21	.27	.56	.03	.89	.37
			Long-delay trials			
Non-updating	.10	.38	.46	.07	.84	.40
Updating	.27	.20	.44	.08	.41	.68
AX	.49	.05	.25	.22	.60	.55
AY	.58	.02	-.21	.26	1.86	.06
BX	.53	.04	.12	.38	1.00	.32
BY	.40	.10	.37	.12	.08	.94

Note. $n = 12$ in each order group.

Discussion

Experiment 5 investigated the recruitment of *g* in the DPX task—a task that was posited to place varying levels of demand on working memory maintenance and updating, and the inhibition of two types of response bias—in the context of a relatively inefficient task conceptualisation (imposed by four-rule task instructions). Patterns of error and speed were very similar to those observed in Experiment 4, in which the only methodological difference was the efficiency of the imposed task conceptualisation (instructions were presented as two rules). Performance on trials that required the inhibition of a prepared prepotent response (AY trials) was significantly worse than performance on other trials (specifically, accuracy on AX trials and response time on all other trial types). One inconsistency between the two experiments was that in Experiment 4 AY trials were also more error-prone than BX trials. That this was not found here suggests that trials that require the inhibition of a prepotent target response indicated by the probe (BX trials) did present some difficulty. However, like Experiment 4, response time was significantly faster on BX, compared to AX, trials which further supports the notion that responses to B trials are faster than A trials because B cues always indicate a non-target response, whereas A cues indicate a target response most, but not all, of the time.

Performance patterns for cues and blocks, to which updating and maintenance demand manipulations were made, respectively, were also similar to those observed in Experiment 4. Accuracy and response time to cues was significantly poorer on trials in which the cue differed from the last cue presented (updating cues) relative to trials in which the cue was constant (non-updating cues); this provides further support for the suggestion that, whether a genuine measure of updating or not, the former type of cue is more demanding than the latter. Cue performance (accuracy and response time) was significantly worse in the block that required maintenance of the cue across a long, compared to a short, delay. As noted in the previous experiment, this could be explained by the shorter inter-trial interval in the long-delay block in which participants

have less time to prepare a response to the following cue. Manipulations of maintenance demand did not affect trial performance in the current experiment which could either suggest that context processing is intact in healthy people or that the long delay length adopted here was not long enough to be demanding for participants. If the latter argument is true, then MacDonald et al.'s (2005) suggestion that the delay length in long-delay trials can be reduced in the DPX task relative to the AX task may be incorrect. Further testing is needed to clarify this. Another consistency with previous experiments was that practice effects were observed. Short–long participants were less accurate (relative to long–short participants) on BX trials in short-delay, and long–short participants were less accurate (relative to short–long participants) on updating cues in long-delay.

Unlike Experiment 4, in which only three measures were significantly correlated with Culture Fair scores, many significant *g* correlations were observed here, and differences in the recruitment of *g* between low/absent and high/present demand conditions were observed. Culture Fair error was significantly correlated with BX and BY trials (accuracy and response time) in short-delay, and all cue and trial types in long-delay (accuracy on each cue type, AY and BX trials; response time on updating cues, AX and BY trials). Correlation with Culture Fair error was significantly higher for accuracy on (a) trials that involved the inhibition of a prepotent but inappropriate response to X probes (BX) relative to non-inhibition trials (AX in short-delay), (b) cues that required updating relative to cues that did not (in short-delay), and (c) non-updating cues in long-delay relative to short-delay. Furthermore, the correlation for response time to BX trials was significantly stronger in Experiment 5 relative to Experiment 4. An unexpected finding was that the correlation with *g* was significantly larger for BY relative to AX trials (in short-delay) which were both considered to be low-demand conditions; perhaps the relatively low abundance of BY trials (5% frequency), and therefore their novelty, increased their association with *g*. Collectively, the correlational findings show that task-recruitment of *g* is strengthened by an increase in real-time

execution demand in the context of a relatively inefficient task conceptualisation. Although there was some degree of overlap in the measures that were associated with *g*, some correlations differed significantly across order groups, which was not the case in Experiment 4. For AY trials, the Culture Fair error correlation was marginally larger in short–long relative to long–short participants for error in short-delay and response time in long-delay.

The efficiency of participants' own task conceptualisations was also associated with Culture Fair performance. Reconceptualisers (participants that selected the one-, two-, or three-rule formats) produced significantly fewer Culture Fair errors than non-reconceptualisers (participants that selected the four-rule format). Furthermore, efficient conceptualisers (participants that selected the one- or two-rule format) made significantly fewer Culture Fair errors than inefficient conceptualisers (participants that selected the three- or four-rule format). Although it has been inferred from previous experiments that two- and four-rule formats reflect efficient and inefficient task conceptualisations, respectively, it is problematic to assume that such linguistic chunking of task rules reflects the actual cognitive representation of task constraints. Using the DPX task as an example, mental "task models" are likely to contain a representation of important stimuli (e.g., the visual characteristics of A cues and X probes), responses (e.g., the spatial location of red and green keys and the speeded responses required), response timings (e.g., temporal locations of cue and pair responses in a trial), trail timings (e.g., short and long delays between the presentations of the cue and the probe), and stimulus–response rules (e.g., respond red to cues and non-target pairs, respond green to target pairs; see Duncan et al., 2008). The strong correlation with Culture Fair error for the number of rules in the selected rule format (r = .76) indicates that the efficiency of the linguistic representation may in some way reflect individual differences in psychometric *g*.

Although, independently of Culture Fair scores, there was a trend for better performance in efficient conceptualisers relative to inefficient conceptualisers for many

measures, the only measure that was performed significantly better by efficient conceptualisers was BX trial accuracy in short-delay. However, the number of rules in the participants' own task conceptualisations was significantly correlated with AY (accuracy in long-delay), BX (accuracy in both blocks and response time in short-delay), and BY (accuracy in short-delay) trials. So, although performance did not generally differ significantly between efficient and inefficient conceptualisers, there was a direct link between number of rules and amount of error/speed of response such that fewer rules reflected better performance. However, because participants selected the rule options post task execution, it is difficult to ascertain with any confidence whether the chosen conceptualisation was applied to the task *during* performance. Participants were instructed to impose their own order on instruction rules prior to completing the task in Experiment 6 in order to more reliably investigate the effects of conceptualisation on performance.

Experiment 5 showed some evidence for the overarching importance of task conceptualisation complexity to the recruitment of *g* relative to the complexity in real-time task execution. The correlation with Culture Fair error for the number of rules in the chosen rule format did not change significantly when controlling for any of the performance measures, suggesting that performance did not drive this correlation. However, some of the correlations between Culture Fair error and performance (response time on AY and BX trials in short-delay) were reduced significantly when controlling for the number of rules in the chosen format, suggesting that the number of rules drove these correlations. These findings do not completely settle the argument that task conceptualisation is more central to *g* than task execution demands because a significant reduction in the strength of performance–*g* correlations was not observed for every performance measure. It is, however, interesting that the two affected performance–*g* correlations were for the two measures that involved inhibition of a prepotent response bias. Therefore, these are the first set of findings that at least help to

clarify the relative importance of inhibition demand and inefficient task conceptualisation to the recruitment of *g*.

In conclusion, the results of Experiment 5 suggest that the recruitment of *g* in task performance is greater when operative task-relevant information is presented as four (Experiment 5), relative to two (Experiment 4), distinct rules. Furthermore, increases in real-time execution demands (through manipulation of inhibition, updating, and maintenance demands) acted to increase correlations with *g*, but only in the context of a relatively inefficient imposed task conceptualisation. Experiment 5 also provides evidence that efficient task reconceptualisation is not only generally restricted to high-*g* individuals but is also perhaps more important to the recruitment of *g* in task execution than prepotent response inhibition requirements. Experiment 6 investigates the task modelling function in more detail. Participants complete the colour shape match and DPX tasks using a task switching paradigm, and impose their own order on instruction rules prior to task execution. Such a method is hoped to more clearly assess the relationship between task performance and participants' own task conceptualisations, and also to enable an exploration of both the relationship between task switching and *g* and the temporal aspects of the task model.

CHAPTER 7

Task Switching and Spearman's *g*

Overview

Chapter 7 explores the temporal components of task modelling by employing, in Experiment 6, the colour shape match and DPX tasks under a task switching framework proposed to enhance external validity. The experimental task required varying levels of real-time execution demand, depending on whether or not a trial required inhibition, updating, maintenance, or task switching. In previous experiments, task conceptualisation complexity was defined by the format of the presentation of task instructions. A new method was adopted in Experiment 6 in which participants actively selected a rule format that best reflected their own mental representation of task-relevant information *prior to* task execution. The results showed that correlations between performance on the experimental task and Culture Fair error were neither increased by manipulations of real-time task demand, nor by the number of rules in the selected rule format. However, significant correlations with the number of rules selected were observed for both Culture Fair error and conditions requiring inhibition demand. Furthermore, participants that chose fewer rules produced significantly fewer Culture Fair errors than participants that chose a greater number of rules, suggesting some link between the ability to form an "efficient" task conceptualisation and *g*. In contrast to previous findings, these results do not indicate the greater involvement of task conceptualisation relative to real-time task execution demand in the recruitment of *g*, at least under the constraints of this task switching paradigm.

Experiment 6

The studies presented in earlier chapters indicate that the recruitment of *g* in task performance depends on the complexity of task conceptualisation (i.e., the form in

which task demands are represented in mind), a position consistent with Duncan et al. (2008). Such "complexity" was implied by the number of distinct rules presented at task instructions, whilst keeping the amount of presented task-relevant information constant. In conditions of high real-time task execution demand, the recruitment of *g* in task performance was significantly greater when task instructions were presented as four (Experiments 2 and 5), relative to two (Experiments 1 and 4), distinct rules. Demand was quantified according to conditions requiring (a) inhibition of prepotent response tendencies, (b) maintenance of information across a delay, and (c) updating of information to be maintained. The recruitment of *g* in task performance was also significantly greater when demand was high relative to low, but only when task instructions were presented as four (but not two) distinct rules. The most compelling evidence for the fundamental importance of task conceptualisation efficiency (over and above real-time task execution demand) to Spearman's *g* was that task conceptualisation efficiency drove the relationship between inhibition and *g*, but the relationship between task conceptualisation and *g* was unchanged when controlling for real-time execution demands (Experiment 5).

One potential limitation with these experiments is that they were observed under single-task frameworks. This questions their generalisability to real-world situations because, in everyday life, as the environment changes, it often requires adapting from the performance of one task to the performance of another. One way to address this challenge to external validity of the findings is to explore the same theoretical issues under a task (switching) framework in which performance of the colour shape match task is interrupted, at unpredictable intervals, by performance of the DPX task, and vice versa. Task switching refers to shifting attention from one task or goal to another (e.g., Monsell, 2003). Traditional task switching paradigms generally employ stimuli that afford two or more tasks; participants are required to switch between the performance of one task and the performance of another. The ability to rapidly and efficiently switch between, for example, classifying numbers as odd or even to classifying numbers as

high or low, is generally measured by performance decrements following a task switch. These *switch costs* are reflected by an increase in reaction time and sometimes error.

Research has shown that switch costs can be amplified by a number of task manipulations, such as switching the task at unpredictable intervals, decreasing the amount of time in which the new task can be prepared, and switching to a task that requires episodic (rather than semantic) retrieval from long-term memory. For example, Monsell, Sumner, and Waters (2003) presented single-digit numbers which required a classification of either high/low or odd/even. A cue, in the form of the background screen colour (pink or blue) or background shape (square or diamond), was presented before each trial which indicated the current task. Switch costs were observed following all task switches, but were reduced when switches were *predictable* (when the task switched every four trials rather than randomly trial by trial) and *prepared* (when the duration of the cue was increased). This was reflected by the recovery of performance by the second trial following a switch when switches were predictable, but not until the third or fourth trial when switches were unpredictable, and by a smaller difference in reaction time between the first and second trial (after the switch) as cue duration increased.

Mayr and Kliegl (2000) were interested in the relationship between task switching and the retrieval of information from long-term memory. In their computer-based task, 16 concrete nouns (which corresponded to small/large living/non-living objects) required classification. Semantic categories (size and living/non-living) were determined by the meaning of the word, and episodic categories (location on the computer screen and font colour) were pre-determined in a learning-to-criterion phase prior to task execution. Switch costs were observed when switching to either type of judgement, but were larger when switching from semantic to episodic judgements. These findings were interpreted on the theoretical grounds that episodic retrieval imposes greater retrieval effort. The noun *mouse* always refers to something that is small and something that is living, but only refers to bottom or top (of the screen) and

yellow or blue due to a predetermined rule. As such, only episodic classification relies on the retrieval of appropriate contextual rules from long-term memory. The authors conclude that the *advance reconfiguration* (Rogers & Monsell, 1995) aspect of task switching, or *shifting set*, is closely related to the retrieval of information from long-term memory.

It is important to note that the present experimental paradigm differs from traditional task switching paradigms; here, performance switches between two tasks that involve different stimuli, instead of between two tasks that involve the same stimuli. Indeed, there are reasons to suspect reduced switch costs in the present study relative to those generally observed in the literature. In one of the first task switching experiments (Jersild, 1927) switch costs were observed when performance switched between two tasks that were conducted on the same stimuli (adding vs. subtracting three from a number) but not when two tasks were conducted on separate stimuli (adding three to a number vs. writing the antonym of an adjective). Jersild (1927) also discovered that the ability to shift set was closely related to performance on standard intelligence tests (in children and college students). The relationship between intelligence and task switching has since been debated, with some studies reporting a strong relationship (e.g., Salthouse et al., 2003) and others reporting non-significant relationships (e.g., Dempster, 1991; Rockstroh & Schweizer, 2001; Friedman et al., 2006).

There were a number of reasons for exploring the relationship between *g* and task switching using the current paradigm in the place of more traditional methods. In order to address the challenge to external validity of previous findings, it was important not to alter the demands of the tasks, or the instructions, other than by introducing the requirement to switch tasks. If a more traditional approach was adopted, the task instructions would need to reflect any new task requirements associated with performing two tasks on the same stimuli. For example, the stimuli in the colour shape match task could have been classified in terms of colour (red, blue, or green) or shape (circle, square, or triangle) depending on the position of the tick (left or right).

Although a task of this kind involves the same stimuli as previous experiments, a new set of task rules would be associated with performance; such a method would have prevented an exploration of the effects of increased external validity on task performance. It may even be argued that the adopted approach is more like real-world experience. Tasks in everyday life are typically interrupted by the requirement to attend to a task with very different constraints and stimuli, such as attending to a phone call when tea-making.

Despite its differences with classic task switching experiments, the proposed task has been designed so that switches are unpredictable and do not incur any preparation time, thus it is conceivable to suspect switch costs. Furthermore, the task sets associated with each task are more episodic than semantic in nature; trials are classified as targets (single-matching trials; AX trials) or non-targets (non- and double-matching trials; AY, BX, and BY trials) on the basis of a novel predetermined rule (outlined in task instructions), rather than a long-standing semantic rule. Jersild (1927) suggested that when the two to-be-completed tasks merge to form an integrated representation of the task, switch costs are reduced. In order to encourage the construction of two separate task conceptualisations which required retrieval from long-term memory to working memory focus (Mayr & Kleigl, 2000) each time the task switched, participants were first instructed that the experiment involved two separate tasks, and were then given (and imposed their own order on) the instructions for each task in turn.

Specifically, sub-blocks of the colour shape match task were performed between sub-blocks of the DPX task. The number of trials in each sub-block varied (between three and 20 trials) such that task switches were specified at unpredictable intervals. Participants imposed their own order on task instructions prior to, instead of post (as in Experiments 3, 4, and 5), task execution; task instructions for each task were initially verbally presented to participants as four distinct rules (identical to the instructions of Experiments 2 and 5) and participants then selected a rule format that best reflected how the task was represented in their own mind. This chosen task conceptualisation (which

comprised one to four rules) was used to split participants into high/low task conceptualisation efficiency groups. As in previous experiments, performance scores and correlations between performance scores and scores on the Culture Fair were compared across conditions that differed in terms of real-time task execution demand as well as across rule groups. However, only descriptive data is presented for rule group comparisons because the sample sizes were very small thus any statistical effects, or non-effects, would have low reliability. Switch costs were also compared across both real-time demand conditions and rule groups in order to address whether retrieval demand is affected by these manipulations.

On the basis of the findings of previous experiments, it was predicted that the correlation with Culture Fair error would be numerically stronger, and that performance would be significantly poorer, in the groups that selected a greater number of rules (relative to a lower number of rules), particularly in high real-time task demand conditions. If the ability to switch task is related to *g* then correlations with Culture Fair error were expected to be significantly stronger on trials in which the task switched (termed *switch trials*) relative to *non-switch trials*.

Method

Participants

An opportunity sample of 48 adults (29 female) aged between 18 and 59 years (M = 21.00, SD = 14.95) was recruited from the Department of Psychology, Anglia Ruskin University and the wider community. Students received credit for their participation. None of the participants had participated in Experiments 1 through 5, disclosed a history of neurological disorder, or were bilingual (due to research suggesting that bilingualism may be associated with enhanced performance in task switching studies; e.g., Prior & MacWhinney, 2010). Participants produced between 3 and 23 Culture Fair errors (M = 11.88, SD = 4.58).

Materials

Colour shape match–DPX task. The colour shape match–DPX task was programmed in EPrime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). The task comprised 120 colour shape match trials and 120 DPX trials. Colour shape match trials were non-, single-, and double-matching pairs of coloured shapes. DPX trials were cue-probe pairs consisting of the consecutive presentation of a fixation cross, a cue, a blank-screen delay, a probe, and a blank-screen inter-trial interval. Sub-blocks containing three to 20 colour shape match trials were presented between sub-blocks containing three to 20 DPX trials.

The 120 colour shape match trials were identical in form and order to the inhibition block of Experiments 1 through 3 (see pages 44–46 and 48–49). That is, in every 10 frames: two of the first five frames were single-matching with at least one non-matching trial in between; the sixth and eighth frames were non-matching; in 75% of trials the seventh frame was double-matching; either the ninth or 10th frame was single-matching and the other was non-matching. Altogether there were 75 non-matching trials (62.5% frequency), 36 single-matching trials (30% frequency), and nine double-matching trials (7.5% frequency).

The 120 DPX trials were almost identical in order to the short-delay block of Experiments 4 and 5 (repeated three times; see pages 123–126 to 127–128). Minor alterations to the order were necessary to ensure that the frequency of each type of trial was consistent with previous experiments. Altogether there were 84 AX trials (70% frequency), 15 AY trials (12.5% frequency), 15 BX trials (12.5% frequency), and six BY trials (5% frequency). In every 40 trials there were 10 inhibition trials (five AY and five BX) and 12 updating trials (six A cues and six B cues).

The task was presented in three blocks. The first block comprised 40 trials: one colour shape match sub-block (comprising 20 trials) and one DPX sub-block (comprising 20 trials: 10 consecutive short-delay trials and 10 consecutive long-delay trials). The second and third blocks each comprised 50 trials (25 colour-shape match

trials and 25 DPX trials) and twelve sub-blocks (six colour shape match sub-blocks and six DPX sub-blocks). Sub-blocks in the second and third blocks comprised a varying number of trials (between three and 20).

Design

A mixed design was adopted in which performance (accuracy and response time) and correlations between performance and Culture Fair error were compared across performance measures (repeated measures: critical error vs. miss vs. false positive vs. criterion fail vs. response time), trial types (repeated measures: AX vs. AY vs. BX vs. BY), cue types (repeated measures: updating cues vs. non-updating cues), delay lengths (repeated measures: short-delay vs. long-delay), and rule groups (independent samples: one-rule vs. two-rule vs. three-rule vs. four-rule). Participants completed one of four versions of the task which differed only in the order of presentation of (a) the colour shape match sub-blocks and the DPX sub-blocks, and (b) the short-delay and long-delay DPX trials. Each version was completed by the same number of participants (n = 12) in order to prevent specific ordering of each type of trial affecting the data. Specifically, each block of versions A and B commenced with a colour shape match sub-block whereas each block of versions C and D commenced with a DPX sub-block. DPX sub-blocks in the first block were a short-delay sub-block followed by a long-delay sub-block (versions A and C) or a long-delay sub-block followed by a short-delay sub-block (versions B and D). DPX sub-blocks in the second block comprised only short-delay trials (versions A and C) or only long-delay trials (versions B and D), and the DPX sub-blocks in the third block comprised only long-delay trials (versions A and C) or short-delay trials (versions B and D). As in previous experiments, the sequencing of trials varied pseudo-randomly to ensure the separation of the different types of demand within each task; this random order was fixed for all participants.

Procedure

Participants were tested individually in a quiet and brightly lit room. The Culture Fair was completed first, exactly as specified in the manual. Participants were then given the instructions for each task in turn, whilst understanding that the experiment would switch between the two tasks at unpredictable intervals. For the colour shape match task, participants were instructed to respond to items that matched in colour or shape by pressing the button corresponding to the placement of the tick, and to ignore items that matched in both colour and shape and items that did not match. For the DPX task, participants were instructed to respond to all cues by pressing the red button (right, key N), and to respond to probes by pressing the green button (left, key B) on AX pairs and by pressing the red button on all other pairs (AY, BX, and AY). Instructions for each task were initially presented to participants as four distinct rules which were identical to the instructions of Experiments 2 and 5 (see pages 67 and 154), then participants were asked to reflect on how the rules were represented in their own minds and selected a rule format, from a rule sheet (see Appendices L and M), that best reflected their own task conceptualisation. Participants had the option of either choosing from the list of five possible rule formats or writing their own description of how the tasks were represented in their mind. See Appendix N for full research protocol.

The monitor was placed approximately .50 m in front of the participant. At the start of the task 'READY? Press the space bar to start' was presented on the screen. When the space bar was pressed the first block commenced (after an interval of 1,500 ms). Following the first and second blocks 'BREAK Press the space bar to continue' was presented on the screen, which disappeared when the participant hit the space bar and, after an interval of 1,500 ms, the next experimental block commenced. Colour shape match trials were presented for 1,200 ms with a 200 ms blank-screen interval between each trial. DPX trials consisted of a cue (i.e., a dot pattern stimulus representing the letter A or B) presented for 1,000 ms, a blank-screen interval of 1,000 ms (short-delay)

or 5,000 ms (long-delay), a probe (i.e., a dot pattern stimulus representing the letter X or Y) presented for 500 ms, and a blank-screen inter-trial interval of 5,000 ms (short-delay) or 1,000 ms (long delay). Other than between blocks, there was no temporal gap between colour shape match and DPX sub-blocks; the first trial of each DPX sub-block was presented immediately after the 200 ms blank-screen interval of the last trial of each colour shape match sub-block, and the first trial of each colour shape match sub-block was presented immediately after the 1,000/5,000 ms inter-trial interval of the last trial of each DPX sub-block. The whole task took approximately 17½ min to complete.

All key press responses were recorded in EPrime. Responses were attributed to a trial if they occurred within 200 ms to 1,200 ms of stimulus onset; responses that occurred < 200 ms after stimulus onset were assumed to reflect either a late response to the previous stimulus or an anticipatory response to the current stimulus. Participants were not informed of these time limits but were instead asked to respond as quickly and as accurately as possible. Both prior to and post task execution, participants stated the rules for each task (in the format that best reflected their own task conceptualisation). If any task constraint was omitted from these statements the rules were repeated by the experimenter (using their chosen format) until all rules were stated correctly.

Results

Colour shape match task

Performance and *g* correlations. Performance scores and Pearson's correlations (one-tailed) between performance and Culture Fair error are displayed in Table 7.1. Repeated measures ANOVA (two-tailed) showed that error differed significantly between performance measures[22], $F(2.18, 102.61) = 22.39$, $p < .001$, $\eta_p^2 = .32$. Post hoc tests (two-tailed) using Bonferroni correction revealed that error was significantly greater for both critical errors and misses relative to both hand errors and false positives

[22] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

(all $p < .001$). For a sample size of 48, the critical value for Pearson's *r* (one-tailed) is ±.24 at the $p < .05$ alpha level. Misses was the only performance measure that was significantly correlated with Culture Fair error, $r(46) = .34$, $p = .009$. Correlations with Culture Fair error for the majority of other measures were negligible. Thus, the correlation with Culture Fair error was not statistically compared across measures. The correlation between age and Culture Fair error was non-significant, $r(46) = .18$, p = .11.

Table 7.1

Performance scores and Pearson's correlations between Culture Fair error and performance for colour shape match trials

	Performance		*g* correlation	
Measure	*M*	*SD*	*r*(46)	*p*
Critical error	35.30	27.49	.06	.34
Hand error	9.00	10.57	-.06	.35
Miss	27.78	20.27	.34	.01
False positive	11.11	18.66	.18	.11
Response time	881	101	-.02	.46

Note. All means are presented as the percentage of total possible error for that measure except for response time which is presented in ms.

Culture Fair z-score charts. Participants were assigned to z-score bins (width of .50 SD) based on their Culture Fair raw error scores. Figures 7.1 and 7.2 display the number of rules selected (from the rule sheet), and performance scores for each measure, across Culture Fair z-score bins. The charts for the number of rules chosen (Figure 7.1) and misses (Figure 7.2) show that participants that scored poorly on the Culture Fair (those scoring ≥ 1 SD below the sample mean) tended to select a rule format containing a greater number of rules (a difference of 1.39 rules), and produced 23% more misses on the colour shape match task, than participants that made fewer Culture Fair errors (those scoring ≥ 1 SD above the sample mean).

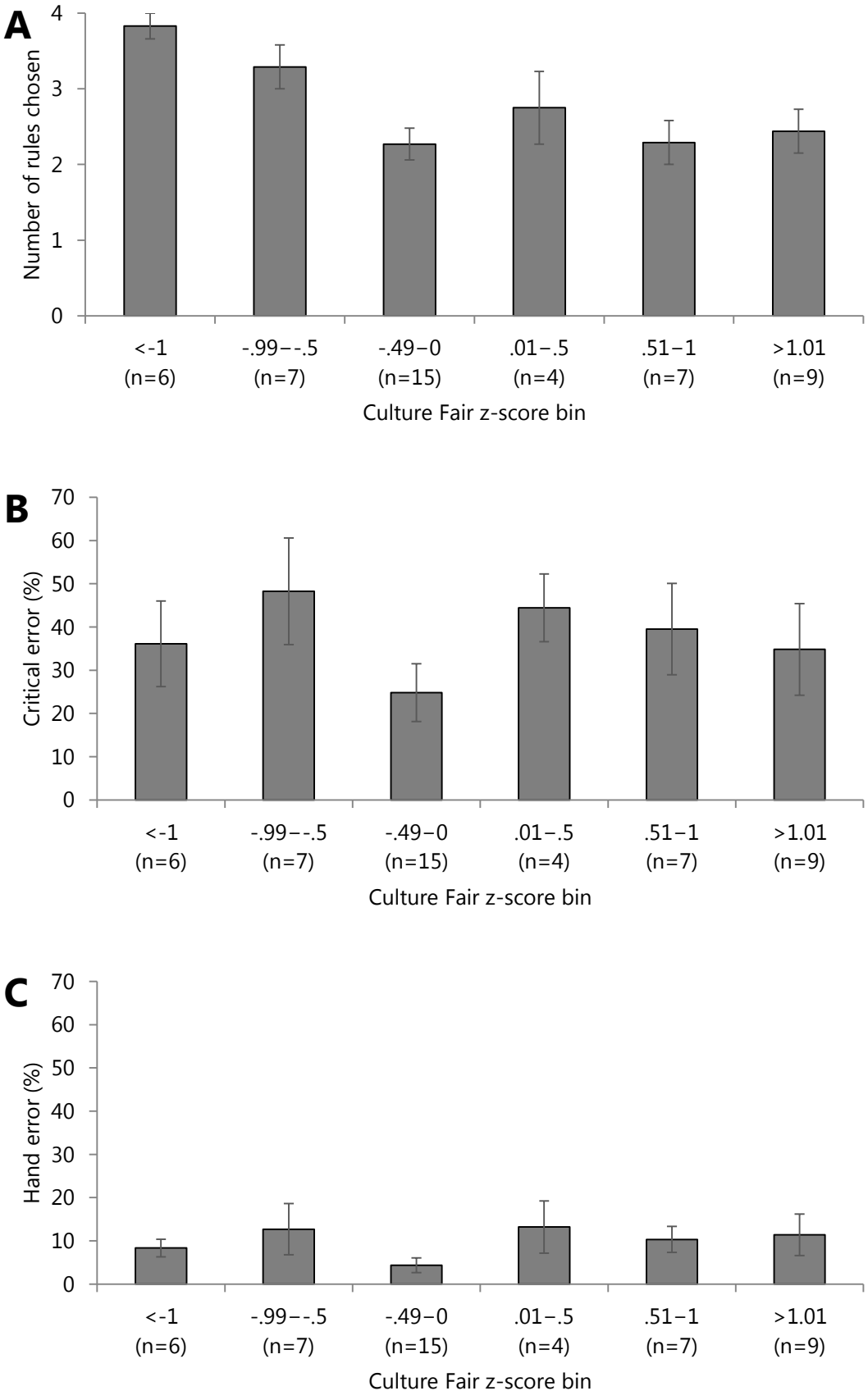


Figure 7.1. Mean number of rules chosen (A) and mean performance (% error) for critical error (B) and hand error (C) across Culture Fair z-score bins. Error bars represent standard error.

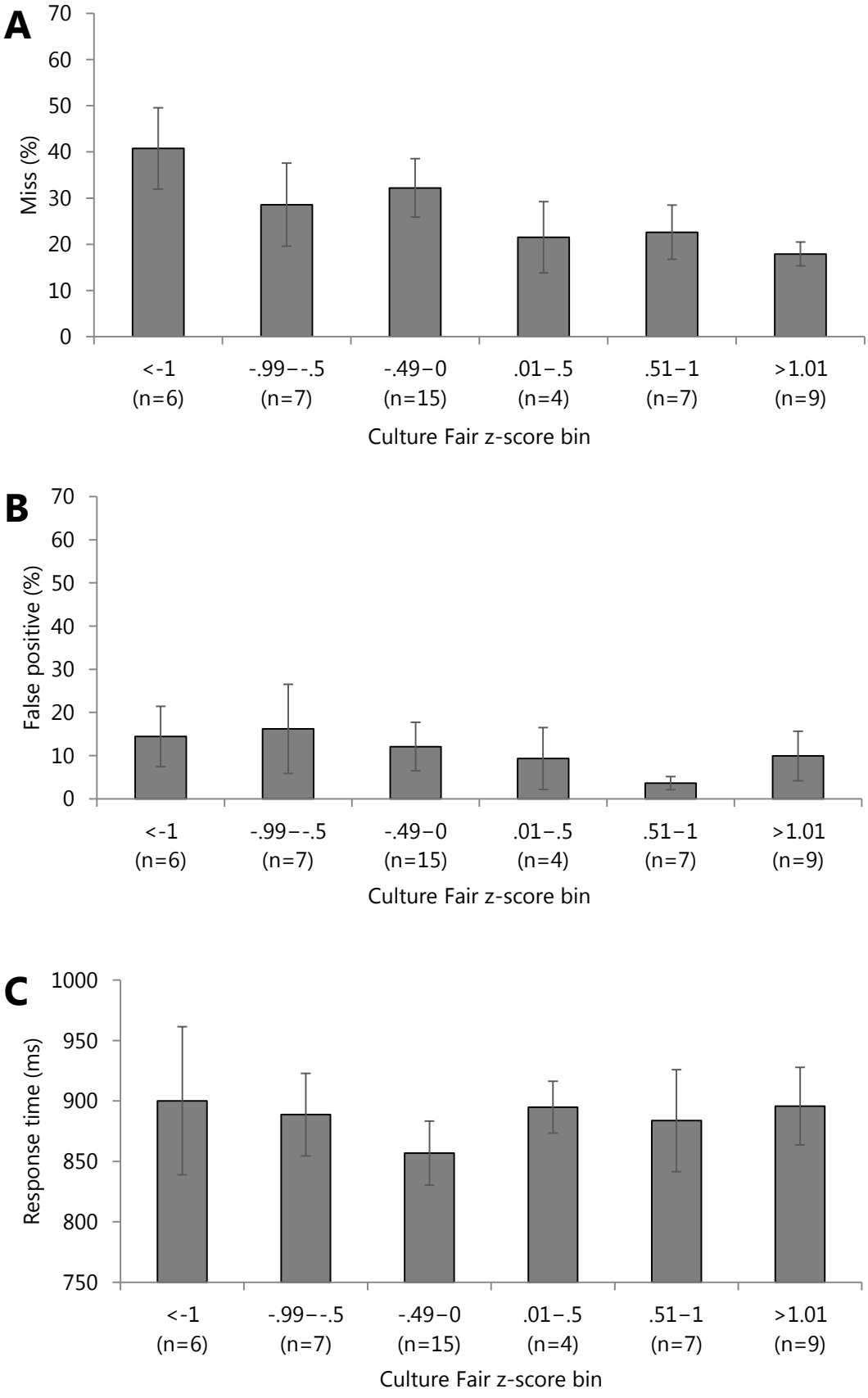


Figure 7.2. Mean performance for miss (% error, A), false positive (% error, B), and response time (ms, C) across Culture Fair z-score bins. Error bars represent standard error.

Task switching. Performance scores and correlations between performance and Culture Fair error (one-tailed), for non-switching trials and switching trials, are presented in Table 7.2. Repeated measures ANOVAs (two-tailed) showed that error was greater on switching trials relative to non-switching trials for misses, $F(1, 47) = 36.23$, $p = .001$, $\eta_p^2 = .44$. However, the opposite pattern of poorer performance on non-switching trials was observed for false positives, $F(1, 47) = 5.32$, $p = .03$, $\eta_p^2 = .10$; response time was also longer on non-switching trials relative to switching trials, $F(1, 47) = 9.27$, $p = .004$, $\eta_p^2 = .17$. Culture Fair error was significantly correlated with: misses on non-switching trials, $r(46) = .32$, $p = .01$; misses on switching trials, $r(46) = .29$, $p = .02$; and critical errors on switching trials, $r(46) = .24$, $p = .05$. William-Hotelling t-tests (two-tailed) revealed that none of the correlations between performance and Culture Fair error differed between non-switching and switching trials (all $p > .10$).

Table 7.2

Performance scores and Pearson's correlations with Culture Fair error for switching and non-switching colour shape match trials

	Performance		*g* correlation	
Measure	*M*	*SD*	*r*(46)	*p*
		Non-switching trials		
Critical error	48.22	25.29	.11	.23
Hand error	9.25	10.75	-.06	.34
Miss	26.10	20.26	.32	.01
False positive	11.72	19.93	.18	.11
Response time	827	103	.001	.50
		Switching trials		
Critical error	41.67	39.05	.24	.05
Hand error	7.29	17.83	.02	.44
Miss	56.25	39.44	.29	.02
False positive	6.02	12.91	.11	.23
Response time	595	496	-.19	.10

Note. All means are presented as percentage error except for response time (ms).

Performance scores and correlations with *g* across rule groups. Participants were split into rule groups based on the number of rules contained in the rule format that they selected at task instructions (for the colour shape task), which ranged between one and four rules. The number of rules selected was significantly correlated with Culture Fair error, $r_s(46) = .46$, $p = .001$. Furthermore, there was a significant difference across rule groups in the number of Culture Fair errors produced, $F(3, 44) = 6.07$, $p = .001$, $\eta_p^2 = .29$. Post hoc pairwise comparisons (two-tailed) showed that Culture Fair error was significantly greater in participants that selected the four-rule format ($n = 10$) relative to participants that selected the one-rule ($n = 5$; $p < .03$, Bonferroni corrected), two-rule ($n = 15$; $p = .001$, Bonferroni corrected), and three-rule ($n = 18$; $p < .01$, Bonferroni corrected) formats. Performance scores for each rule group are presented in Figure 7.3.

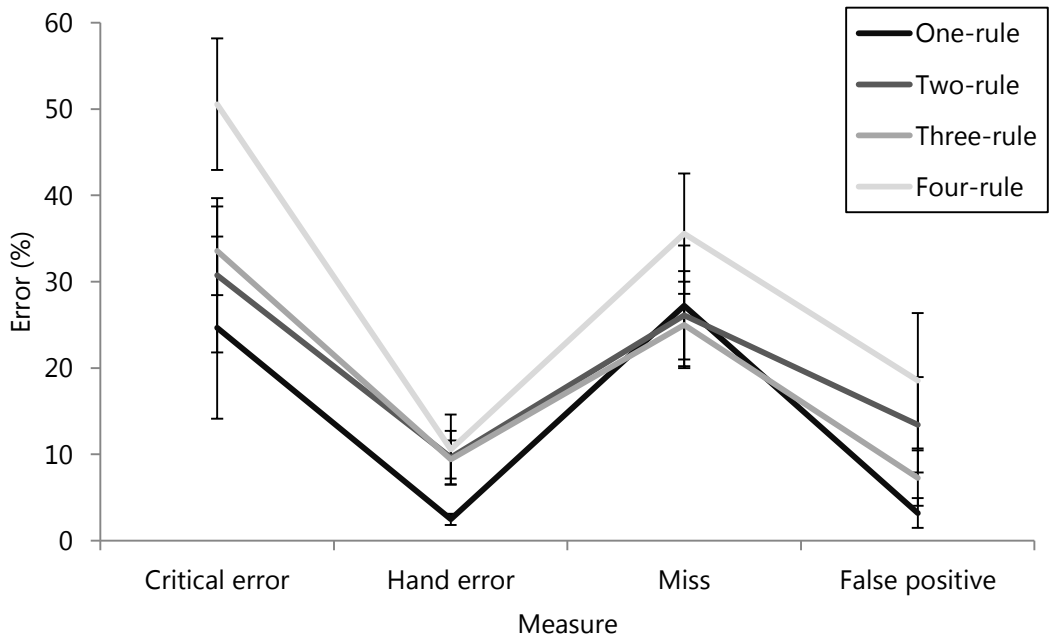


Figure 7.3. Performance scores across rule groups for colour shape match trials. Error bars represent standard error.

Figure 7.3 suggests better performance in participants that selected the one-rule format, and worse performance in participants that selected the four-rule format. However, one-way ANOVAs (two-tailed) showed that performance scores were statistically equivalent across rule groups for all performance measures (all $p > .10$). However, the reliability of comparisons across rule groups is severely challenged due to very small sample sizes, particularly for one- and four-rule groups. Spearman's rank order correlations (one-tailed) showed that the number of rules selected was significantly related to critical error, $r_s(46) = .29$, $p = .04$; all other correlations between the number of rules selected and performance were non-significant (all $p > .10$).

Culture Fair error was correlated with performance in each rule group, with the exception of one-rule due to the small sample size of $n = 5$. Indeed, the reliability of these correlations is reduced because the size of the rule groups was small; the effect of, for example, outliers, transformations of the variables, or dissimilar distribution shapes, is stronger for correlations that are calculated on a small set of data (Goodwin & Leech, 2006). The only significant observed correlation with Culture Fair error was for misses in participants that selected the two-rule format, $r(13) = .57$, $p = .01$; all other correlations were non-significant (all $p > .09$).

Complexity in task execution versus task conceptualisation. As previously noted, Culture Fair error was significantly correlated with misses, $r(46) = .34$, $p = .009$; Culture Fair error was also significantly with the number of rules selected, $r_s(46) = .46$, $p = .001$. A series of partial correlations were performed between Culture Fair error and performance scores (when controlling for the number of rules) and between Culture Fair error and the number of rules selected (when separately controlling for each performance measure). None of these correlations differed in strength from their bivariate equivalent (all $p > .10$) which suggests that the miss–*g* correlation is not driven by the number of rules held in mind, and that the number of rules–*g* correlation is not driven by real-time task demand.

DPX task

Performance. Error (%) and response time (ms) scores for cue types (updating, non-updating) and trial types (inhibition: AY and BX; non-inhibition: AX and BY) across delay lengths (short-delay, long-delay) are presented in Table 7.3. Performance was compared across trial type, cue type, and delay length.

Table 7.3

Performance scores for DPX trials

	Trial type			
	Short-delay		Long-delay	
Measure	*M*	*SD*	*M*	*SD*
		Error (%)		
Non-updating	14.09	23.01	14.93	21.41
Updating	17.36	26.30	16.85	23.47
AX	15.77	21.22	17.91	20.97
AY	38.91	28.25	37.17	29.74
BX	26.64	29.08	29.87	32.69
BY	22.92	34.49	21.53	33.33
		Response time (ms)		
Non-updating	649	167	679	155
Updating	675	131	654	180
AX	682	135	679	137
AY	758	247	732	266
BX	576	281	500	217
BY	557	284	495	255

For error scores, a 4 (trial) × 2 (delay) repeated measures ANOVA (two-tailed) revealed a significant main effect of trial type[23], $F(2.32, 108.79) = 11.93$, $p < .001$, $\eta_p^2 = .2$. However, there was no main effect of delay length, $F(3, 47) = .14$, $p = .71$, $\eta_p^2 = .003$. There was also no trial type × delay length interaction[24], $F(2.51, 117.75) = .73$, $p = .53$, $\eta_p^2 = .02$. Post hoc pairwise comparisons (two-tailed) showed that error was significantly greater on inhibition trials relative to non-inhibition trials: error was

[23] Degrees of freedom and *p* adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

greater on AY trials (M = 38.04, SE = 3.77) relative to AX trials (M = 16.84, SE = 2.93) and BY trials (M = 22.22, SE = 4.69; both $p < .001$, Bonferroni corrected), and error was greater on BX trials (M = 28.26, SE = 4.16) relative to AX trials ($p < .05$, Bonferroni corrected). A 2 (cue) × 2 (delay) repeated measures ANOVA revealed no main effect of cue type, $F(1, 47) = 1.81$, $p = .19$, $\eta_p^2 = .04$. There was no main effect of delay length, $F(1, 47) = .01$, $p = .92$, $\eta_p^2 = .00$. Neither was there a cue type × delay length interaction, $F(1, 47) = .3$, $p = .59$, $\eta_p^2 = 006$.

For response time scores, a 4 × 2 repeated measures ANOVA (two-tailed) revealed a significant main effect of trial type[24], $F(2.33, 109.56) = 21.42$, $p < .001$, $\eta_p^2 = .31$. However, there was no main effect of delay length, $F(1, 47) = 2.78$, $p = .10$, $\eta_p^2 = .06$. There was also no trial type × delay length interaction[24], $F(2.04, 95.66) = .68$, $p = .57$, $\eta_p^2 = .01$. Post hoc pairwise comparisons (two-tailed) showed that response time was slower on A-cue trials relative to B-cue trials: response time was significantly slower on AX trials (M = 680.73, SE = 16.95) relative to BX trials (M = 538.27, SE = 25.42) and BY trials (M = 526.04, SE = 33.18), and on AY trials (M = 745.16, SE = 26.84) relative to BX trials and BY trials (all $p < .001$, Bonferroni corrected). A 2 × 2 repeated measures ANOVA (two-tailed) revealed no significant main effect of cue type, $F(1, 47) = .001$, $p = .99$, $\eta_p^2 = .00$. Neither was there a main effect of delay length, $F(1, 47) = .06$, $p = .80$, $\eta_p^2 = .001$. There was also no cue type × delay length interaction, $F(1, 47) = 1.37$, $p = .25$, $\eta_p^2 = 03$.

***g* correlations.** Pearson's correlations (one-tailed) between performance and Culture Fair error are displayed in Table 7.4. For error scores, correlations with Culture Fair error were significant for updating cues (in long-delay only), and all trial types (AX, AY, BX [in long-delay only], and BY [in both blocks]); significant correlations ranged between $r = .24$ and $r = .36$. Williams-Hotelling t-tests (two-tailed) revealed that none of the correlations with Culture Fair error differed significantly across error for cue type, trial type, or delay length (all $p > .10$). For response time scores, correlations

[24] Degrees of freedom and p adjusted using the Greenhouse-Geisser epsilon procedure due to violation of the assumption of sphericity across measures.

with Culture Fair error were significant for AX trials (in long-delay only) and, although negative, AY trials (in long-delay only) and BY trials (in short-delay only); positive and negative significant correlations were weak at $r = \pm.26$ to $r = \pm.27$. Williams-Hotelling t-tests (two-tailed) revealed that, across trial types in the short-delay block, the correlation with Culture Fair error was significantly stronger for response time to AX trials relative to response time to all other trial types: AY trials, $t(45) = 2.30$, $p = .03$; BX trials, $t(45) = 2.09$, $p = .04$; and BY trials, $t(45) = 2.79$, $p = .008$. The correlation with Culture Fair error was also significantly larger for AX trials relative to AY trials in the long-delay block, $t(45) = 2.61$, $p = .01$. Across delay lengths, the correlation with Culture Fair error for response time to BY trials was significantly stronger in long-delay relative to short-delay (in which it was negative), $t(45) = 2.81$, $p = .007$. All other comparisons were non-significant (all $p > .10$) except the difference between the correlation between Culture Fair error and response time to AY trials relative to BY trials in long-delay which was marginal, $t(45) = 1.94$, $p = .06$.

Table 7.4

Pearson's correlations between Culture Fair error and performance DPX trials

	Trial type			
	Short-delay		Long-delay	
Measure	*r*(46)	*p*	*r*(46)	*p*
		Error (%)		
Non-updating	.17	.12	.13	.19
Updating	.22	.07	.24	< .05
AX	.26	.04	.32	.01
AY	.33	.01	.35	.008
BX	.21	.08	.29	.03
BY	.36	.005	.35	.008
		Response time (ms)		
Non-updating	-.03	.43	-.06	.33
Updating	.19	.09	-.05	.38
AX	.22	.07	.26	.04
AY	-.22	.44	-.27	.03
BX	-.07	.31	.06	.34
BY	-.26	.04	.14	.17

Culture Fair z-score charts. Participants were assigned to z-score bins (width of .50 SD) based on their Culture Fair raw error scores. Performance scores for each measure across Culture Fair z-score bins are presented Figures 7.4 to 7.8. The data show a trend towards better performance, as displayed by both fewer errors and faster response speed, in participants at the higher end of the *g* distribution. Participants that scored ≥ 1 SD above the sample mean on the Culture Fair selected a rule format containing fewer rules (a difference of 1.72 rules) than participants that scored ≥ 1 SD below the sample mean. Participants that scored ≥ 1 SD above the sample mean also produced 13% fewer updating cue errors in long-delay, produced 9%, 19%, and 37% fewer errors on AX, AY, and BY trials (respectively) in short-delay, and produced 16%, 18%, 15%, and 30% fewer errors on AX, AY, BX, and BY trials (respectively) in long-delay, and responded 91 ms slower on AX trials in long-delay, than participants that

scored ≥ 1 SD below the sample mean[25]. However, participants that scored ≥ 1 SD below the sample mean on the Culture Fair actually responded (a) 174 ms *faster* on AY trials in long-delay, and (b) 102 ms *faster* on BY trials in short-delay, than participants that scored ≥ 1 SD above the sample mean.

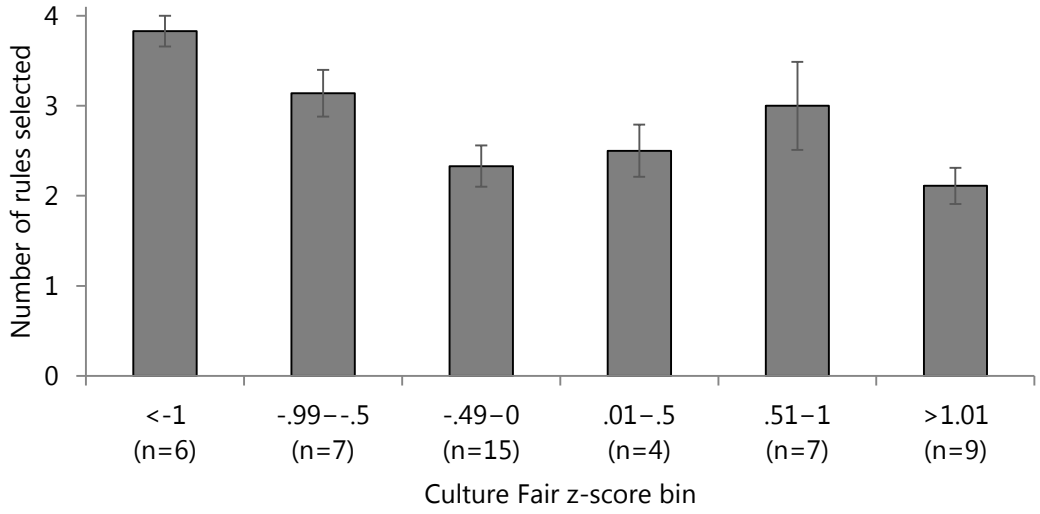


Figure 7.4. Mean number of rules selected for the colour shape match task across Culture Fair z-score bins. Error bars represent standard error.

[25] Note that differences are described for significant correlations only.

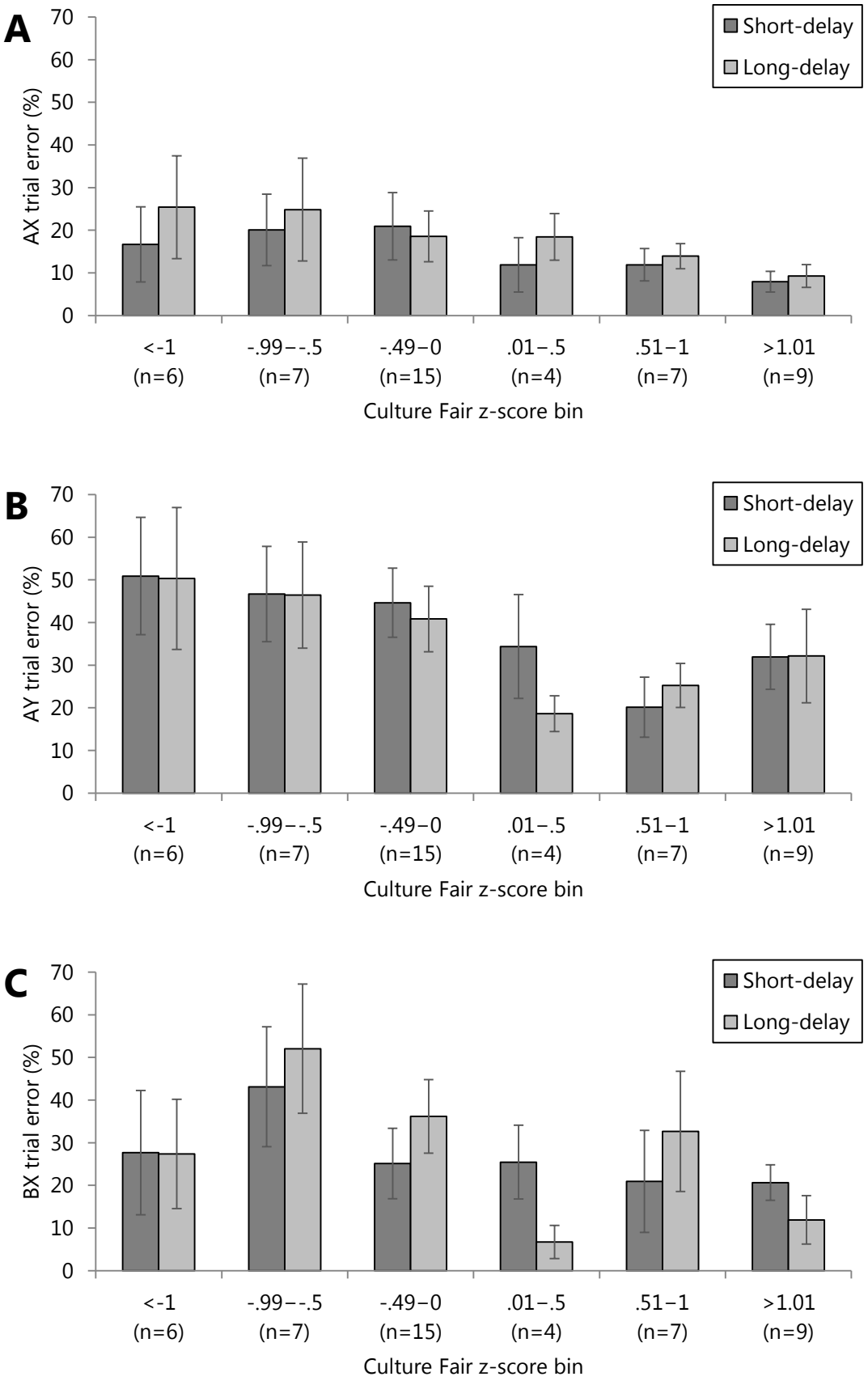


Figure 7.5. Mean error (%) across Culture Fair z-score bins for AX trials (A), AY trials (B), and BX trials (C). Error bars represent standard error.

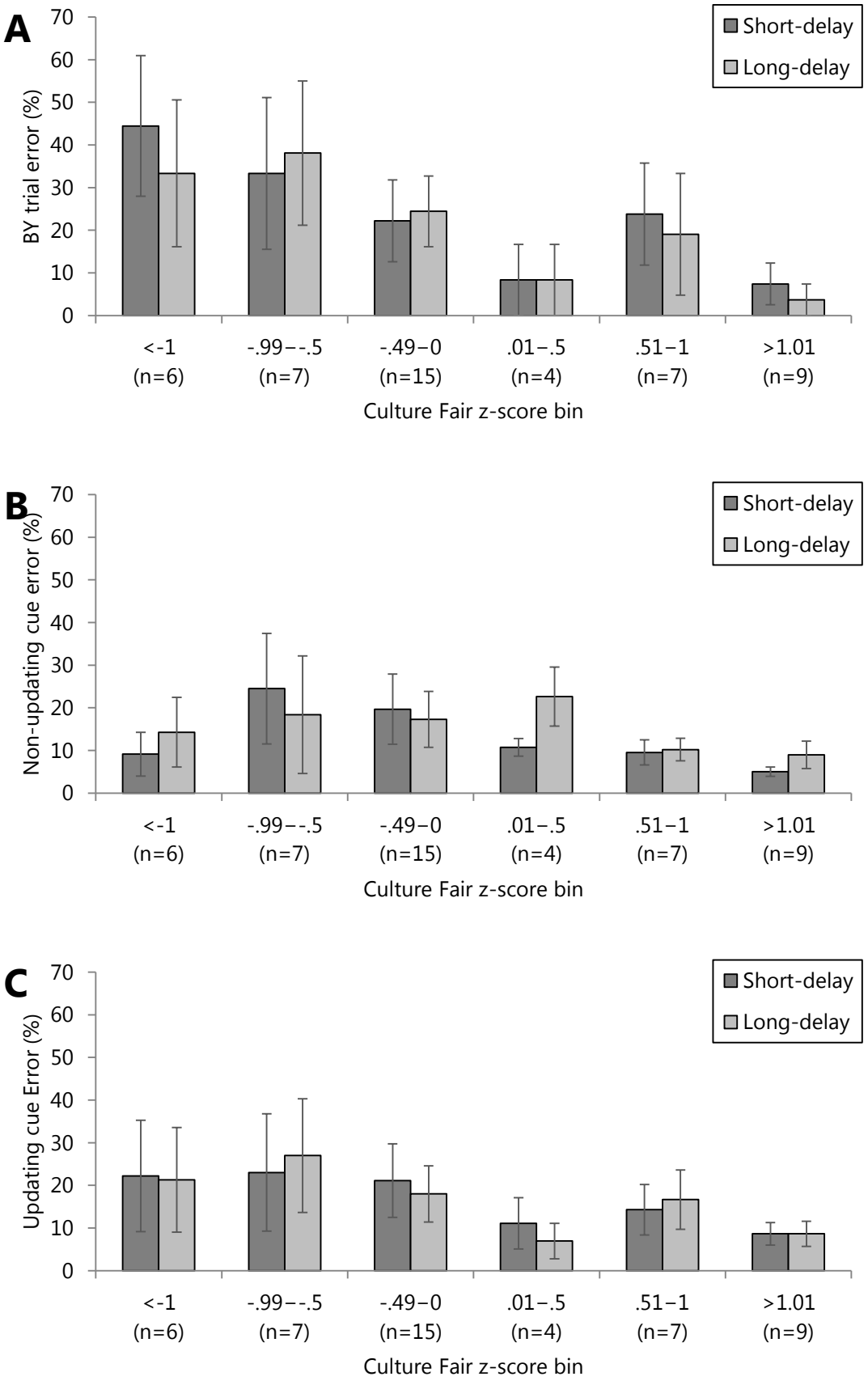


Figure 7.6. Mean error (%) across Culture Fair z-score bins for BY trials (A), non-updating cues (B), and updating cues (C). Error bars represent standard error.

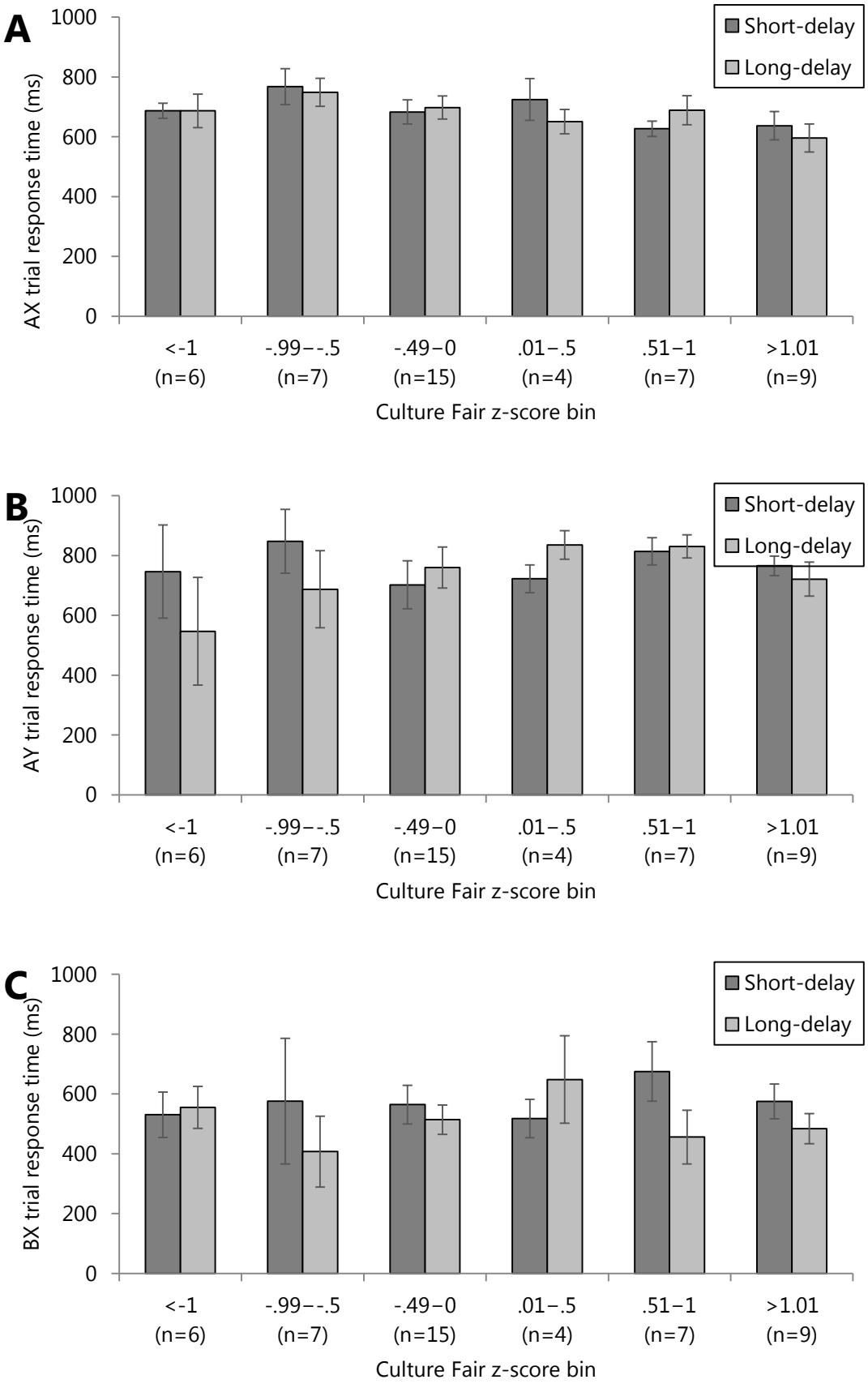


Figure 7.7. Mean response time (ms) across Culture Fair z-score bins for AX trials (A), AY trials (B), and BX trials (C). Error bars represent standard error.

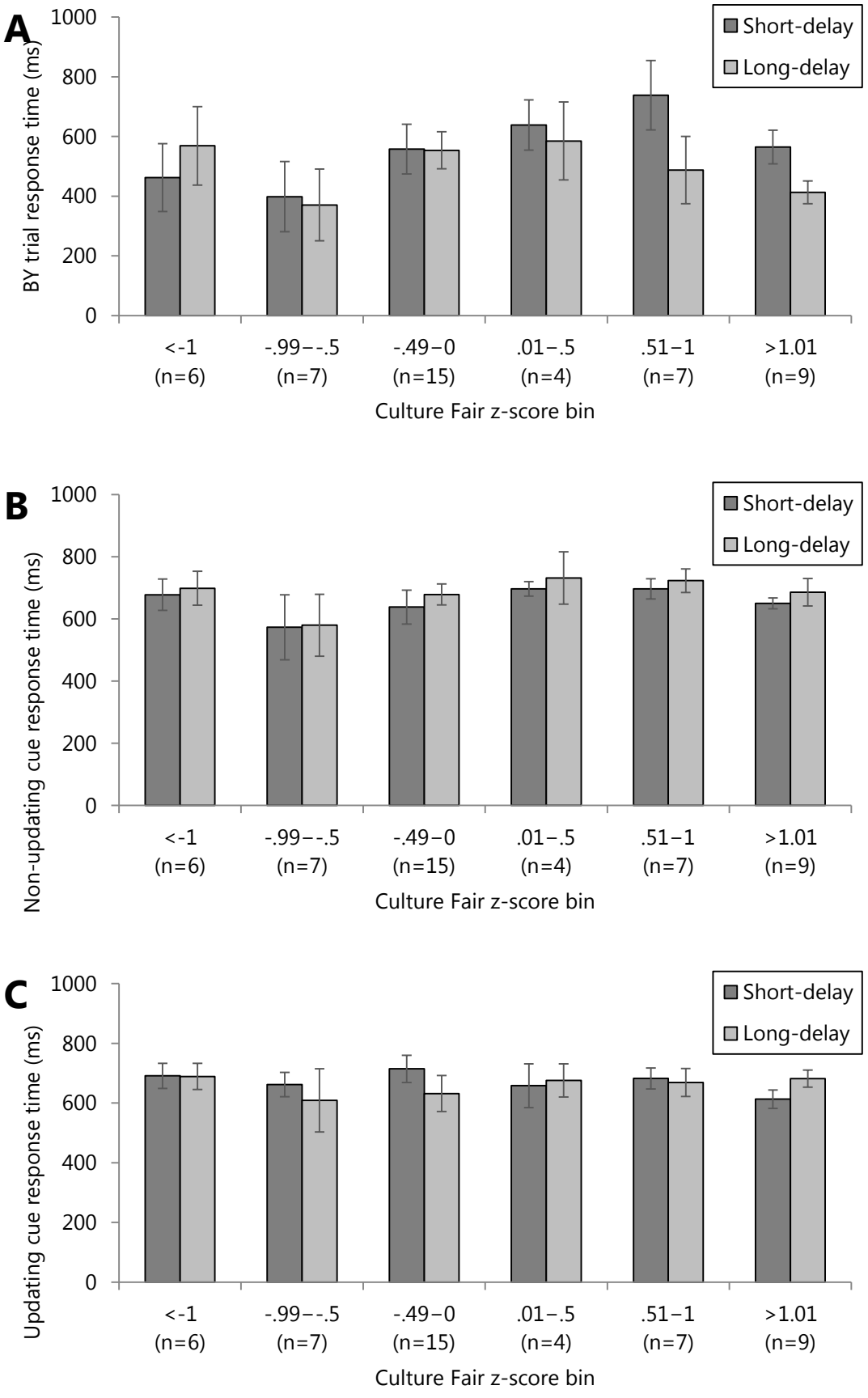


Figure 7.8. Mean response time (ms) across Culture Fair z-score bins for BY trials (A), non-updating cues (B), and updating cues (C). Error bars represent standard error.

Task switching. Performance scores and correlations with Culture Fair error (one-tailed), for non-switching trials and switching trials, are presented in Tables 7.5 (error) and 7.6 (response time). Repeated measures ANOVAs (two-tailed) showed that error was greater on switching trials relative to non-switching trials for: non-updating cues in short-delay, $F(1, 47) = 40.45$, $p = .001$, $\eta_p^2 = .46$; non-updating cues in long-delay, $F(1, 47) = 31.57$, $p = .001$, $\eta_p^2 = .4$; and AX trials in short-delay, $F(1, 47) = 4.90$, $p = .03$, $\eta_p^2 = .09$. Response time was significantly slower on switching trials relative to non-switching trials for: non-updating cues in short-delay, $F(1, 47) = 46.47$, $p = .001$, $\eta_p^2 = .50$; and non-updating cues in long-delay, $F(1, 47) = 10.10$, $p = .003$, $\eta_p^2 = .18$. The opposite pattern of slower response time on non-switching trials was found for: AY trials in short-delay, $F(1, 47) = 3.46$, $p = .001$, $\eta_p^2 = .44$; AY trials in long-delay, $F(1, 47) = 17.36$, $p = .001$, $\eta_p^2 = .27$; and BX trials in short-delay, $F(1, 47) = 12.11$, $p = .001$, $\eta_p^2 = .21$.

Culture Fair error was significantly correlated with error on AX trials (in both blocks), AY trials (in short-delay), BX trials (in long-delay), and BY trials (in both blocks) on non-switching trials, and with AY trials (in short-delay), and BX trials (in long-delay) on switching trials. Significant correlations ranged between $r = .26$ and $r = .36$. For response time, Culture Fair error was significantly correlated with AX trials in both blocks for non-switching trials, and was significant but negative for BX trials in both blocks for switching trials: positive correlations (for AX trials) were each $r(46) = .31$, $p = .02$; and negative correlations (for BX trials) ranged between $r = -.31$ and $r = -.25$. William-Hotelling t-tests (two-tailed) revealed that none of the correlations with Culture Fair error differed across non-switching and switching trials (all $p > .10$). However, for response time, the correlation with Culture Fair error for BX trials (in long-delay) was significantly stronger (but negative) on switching relative to non-switching trials, $t(45) = 2.01$, $p = .05$; all other comparisons were not significant ($p > .10$).

Table 7.5

Error scores and Pearson's correlations with Culture Fair error for switching and non-switching DPX trials

	% error		*g* correlation	
Measure	*M*	*SD*	*r*(46)	*p*
		Non-switching trials		
		Short-delay block		
Non-updating	10.40	23.40	.15	.16
Updating	52.79	239.43	.10	.26
AX	14.91	22.10	.27	.03
AY	37.80	29.58	.31	.02
BX	26.89	30.81	.20	.09
BY	22.92	34.49	.36	.005
		Long-delay block		
Non-updating	12.08	21.53	.15	.16
Updating	52.90	239.31	.09	.26
AX	17.39	21.60	.33	.01
AY	36.91	29.68	.21	.08
BX	29.27	33.79	.26	.04
BY	21.53	33.33	.35	.008
		Switching trials		
		Short-delay block		
Non-updating	33.70	31.43	.15	.15
Updating	-	-	-	-
AX	24.90	30.65	.06	.35
AY	34.38	45.14	.36	.007
BX	37.50	48.92	.19	.10
BY	-	-	-	-
		Long-delay block		
Non-updating	32.35	31.00	.04	.39
Updating	-	-	-	-
AX	22.43	27.81	.17	.13
AY	27.08	44.91	.20	.08
BX	33.33	47.64	.27	.03
BY	-	-	-	-

Table 7.6

Response time scores and Pearson's correlations with Culture Fair error for switching and non-switching DPX trials

	Response time		*g* correlation	
Measure	*M*	*SD*	*r*(46)	*p*
		Non-switching trials		
		Short-delay block		
Non-updating	530	145	.03	.43
Updating	658	164	.05	.37
AX	669	124	.31	.02
AY	776	227	-.03	.42
BX	560	233	-.07	.32
BY	574	324	-.16	.14
		Long-delay block		
Non-updating	599	1401	-.04	.38
Updating	654	180	-.05	.38
AX	667	133	.31	.02
AY	697	279	-.18	.13
BX	476	201	.17	.13
BY	495	255	.14	.17
		Switching trials		
		Short-delay block		
Non-updating	745	257	-.08	.31
Updating	-	-	-	-
AX	6134	221	.13	.20
AY	359	430	-.09	.28
BX	348	337	-.31	.02
BY	-	-	-	-
		Long-delay block		
Non-updating	718	277	-.09	.26
Updating	-	-	-	-
AX	649	241	.04	.41
AY	395	438	-.19	.10
BX	382	352	-.25	< .05
BY	-	-	-	-

Performance scores and correlations with *g* across rule groups. Participants were split into rule groups based on the number of rules contained in the rule format that they selected at task instructions (for the DPX task), which ranged between one and four rules. Similar to the colour shape match task findings, only five participants chose the one-rule format which severely challenges the reliability of these rule group findings; the number of people choosing the other rule formats was again slightly larger (n = 16 for two-rule; n = 15 for in three-rule; and n = 12 in four-rule). The number of rules selected was significantly correlated with Culture Fair error, $r_s(46) = .42$, $p = .003$. Furthermore, there was a significant difference across rule groups in the number of Culture Fair errors produced, $F(3, 44) = 4.30$, $p = .01$, $\eta_p^2 = .23$. Post hoc pairwise comparisons showed that Culture Fair error was significantly greater in participants that selected the four-rule format relative to participants that selected the two-rule format (p = .009, Bonferroni corrected). Performance scores for each rule group are presented in Figures 7.9 and 7.10.

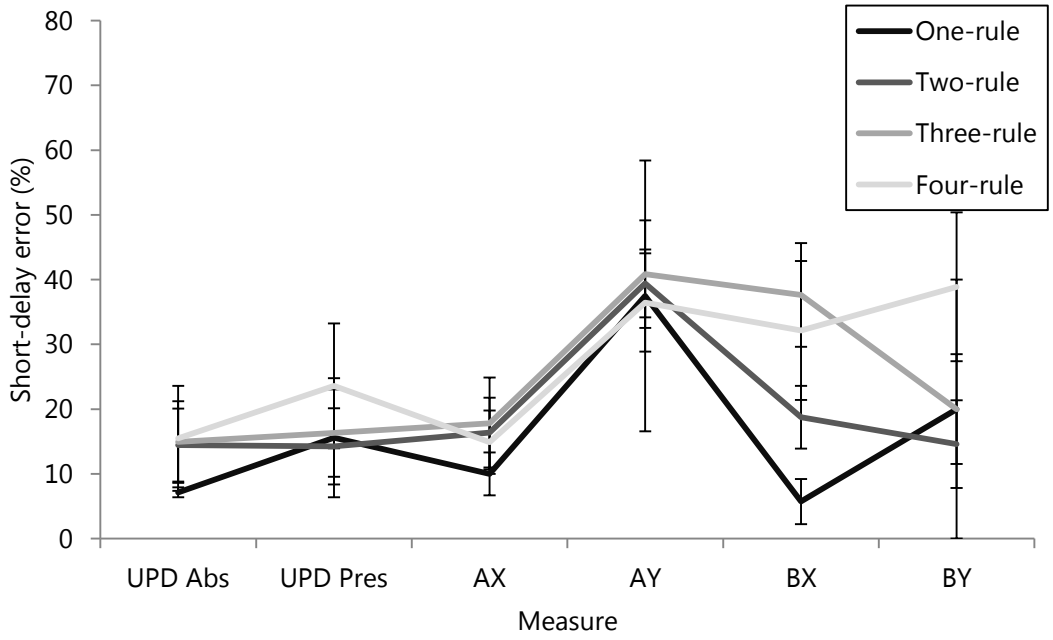


Figure 7.9. Error scores across rule groups in short-delay DPX trials. Error bars represent standard error.

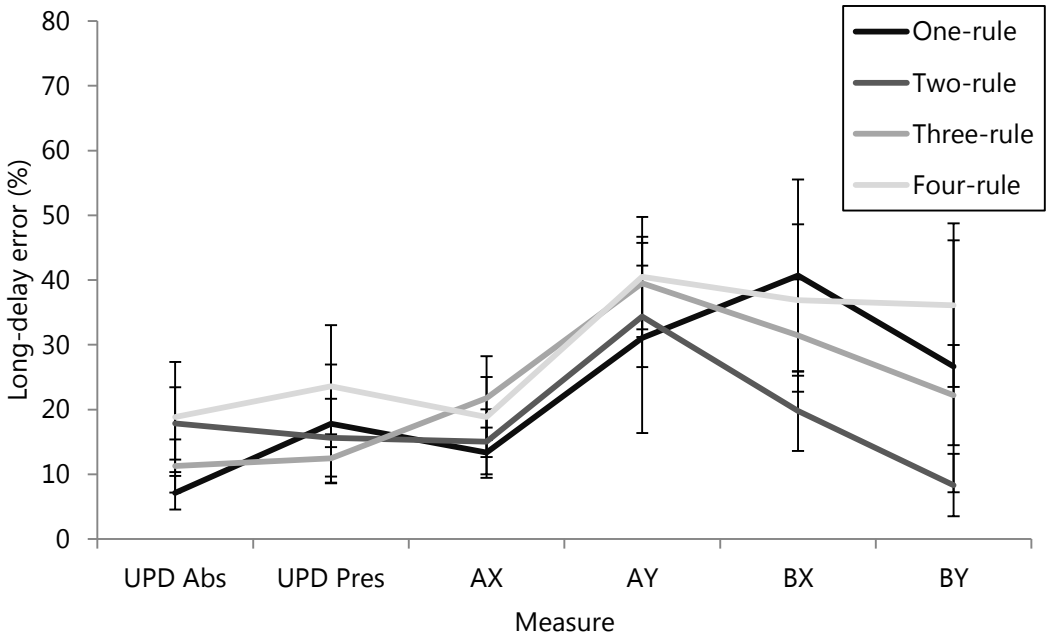


Figure 7.10. Error scores across rule groups in long-delay DPX trials. Error bars represent standard error.

Each of the graphs (Figure 7.9 and Figure 7.10) show trends for better performance in participants that selected fewer rules, but these trends are more noticeable in the short-delay block. One-way ANOVAs (two-tailed), however, showed that performance scores were statistically equivalent across rule groups for all performance measures (all $p > .08$). Spearman's rank order correlations (one-tailed) showed that the number of rules selected was significantly (but negatively) related to response time for BX trials in long-delay, $r_s(46) = -.28$, $p = .05$; all other correlations between the number of rules selected and performance were non-significant (all $p > .10$).

Culture Fair error was correlated with performance in each rule group, with the exception of one-rule due to the small sample size of $n = 5$. However, the reliability of these correlations is reduced because of the small group sizes. In the two-rule group, correlations with Culture Fair error were significant for error on: non-updating cues in short-delay, $r(46) = .45$, $p = .04$; AX trials in short-delay, $r(46) = .48$, $p = .03$; AX trials in long-delay, $r(46) = .48$, $p = .03$; BX trials in long-delay, $r(46) = .58$, $p = .01$; and BY trials in long-delay, $r(46) = .54$, $p = .02$. For response time scores in the two-rule group

correlations with Culture Fair error were significant for: updating cues in short-delay, $r(46) = .65$, $p = .003$; AX trials in long-delay, $r(46) = .74$, $p = .001$; and BY trials in long-delay, $r(46) = .63$, $p = .004$. In the three-rule group, correlations with Culture Fair error were significant for error on: updating cues in long-delay, $r(46) = .51$, $p = .03$; AY trials in short-delay, $r(46) = .46$, $p = .04$; BX trials in short-delay, $r(46) = .46$, $p = .04$; and BX trials in long-delay, $r(46) = .56$, $p = .02$. In the four-rule group, the correlation with Culture Fair error for response time on BY trials in short-delay was significant (but negative), $r(46) = -.51$, $p = .04$.

Complexity in task execution versus task conceptualisation. As previously noted, Culture Fair error was significantly correlated with error on updating cues (in long-delay) and all trials types (AX, AY, BX [in long-delay], BY [in both blocks]), response time on AX trials (in long-delay), and was negatively and significantly correlated with AY trials (in long-delay) and BY trials (in short-delay). The number of rules in the selected format was also significantly correlated with Culture Fair error, $r_s(46) = .42$, $p = .003$. A series of partial correlations were performed between Culture Fair error and performance scores (when controlling for the number of rules in the selected format) and between Culture Fair error and the number of rules in the selected format (when controlling for the each performance measure). None of these correlations differed in strength from their bivariate equivalent (all $p > .10$).

Discussion

Experiment 6 explored: (a) the relative contributions of task conceptualisation versus real-time task execution demand to the recruitment of *g* in task performance under arguably more externally valid conditions (i.e., two experimental tasks were performed, and the performance of each task was interrupted by performance of the other task); (b) whether manipulation of real-time demand (i.e., response inhibition, working memory updating, and working memory maintenance), which affected both performance and task-sensitivity to *g*, extended to task switching; and (c) the temporal

components of the task model.

Manipulations of inhibition demand resulted in poorer performance on high, relative to low, demand conditions, but manipulations of other types of demand did not affect task performance. In line with previous findings, accuracy was poorer on (a) colour shape match trials that required inhibition of two prepotent but inappropriate response biases (critical errors) relative to non-inhibition measures (hand errors and false positives), (b) DPX trials that required inhibition of a prepotent but inappropriate prepared response facilitated by the cue (AY trials) relative to non-inhibition measures (AX and BY trials), and (c) DPX trials that required inhibition of a prepotent but inappropriate response to the probe (BX trials) relative to a non-inhibition measure (AX trials).

Manipulations of working memory updating and working memory maintenance demands, however, did not affect performance levels, and the results for task switching were mixed. Accuracy and response time was statistically equivalent across cues that were assumed to require updating of the cue representation (trials in which the cue differed from the cue in the previous trial) and non-updating cues (trials in which the cue was the same as the cue in the previous trial). Accuracy and response time was also statistically equivalent across trials that were assumed to require maintenance of cue information across a long delay (trials in which the delay between cue and probe presentations was long) and non-maintenance trials (trials in which the delay was short). For task switching, some items were more difficult on switching trials relative to non-switching trials (misses and response time; non-updating cues [accuracy and response time] and AX trials [accuracy in short-delay]), whereas other measures were relatively more difficult on non-switching trials (false positives and response time [colour shape match task]; AY [response time] and BX trials [response time in short-delay]). The task switching nature of the experiment may explain some of the performance findings. A closer inspection of the data revealed that the difficulty of cues increased in Experiment

6 relative to previous experiments (with an average increase of 7–13% for error and an increase of 18–71 ms for speed). Indeed, Monsell (2003) argues that greater working memory load is imposed by blocks in which the task alternates relative to blocks in which the task is repeated because the former requires "keep[ing] track of the task sequence and maintain[ing] two tasks in a state of readiness" (p. 135). The increased difficulty of (at least) cue performance, due to performance interruptions, could also explain why cue performance was unaffected by delay length manipulations in Experiment 6, but in previous experiments cue accuracy and response time was poorer on long-delay trials relative to short-delay trials.

The correlational findings provided limited evidence for manipulations of response inhibition, updating, maintenance, and task switching demands affecting the involvement of *g* in task performance. In support of the importance of updating demand to *g*, significant *g* correlations were observed for error on updating cues (on long-delay trials) but not non-updating cues. For inhibition, maintenance, and switching measures, however, significant correlations were observed for both high (inhibition: AY and BX trials [accuracy]; maintenance: updating cues and AX, AY, BX, and BY trials in long-delay [accuracy]; switching: misses, critical errors, AY trials in short-delay [accuracy], BX trials in long-delay [accuracy], BX trials [response time] on switching trials) and low (non-inhibition: AX trials and BY trials [accuracy], misses; non-maintenance: AX, AY, BX, and BY trials [accuracy] in short-delay; non-switching: misses, AX and BY trials [accuracy], AY in short-delay [accuracy], BX trials in long-delay [accuracy], AX trials [response time] on non-switching trials) demand conditions. Furthermore, critical errors were not significantly correlated with *g*, and the significant (but weak) inhibition–*g* correlation for response time on an inhibition trial (AY in long-delay) was in the opposite direction to that expected; people with higher *g* actually responded *slower* on AY trials in long-delay than people with lower *g* (although the correlation with *g* for AY trial error in long-delay was in the expected direction).

Direct statistical comparisons of *g* correlations across high demand and low demand conditions were also unsupportive of the importance of real-time task execution demand to *g*. For inhibition, *g* correlations were actually stronger for response time on a non-inhibition measure (AX trials) relative to inhibition measures (AY trials, BX trials in short-delay), which is opposite to what would be expected if inhibition demand acted to increase correlations with *g*. However, the response time findings should be taken with caution due to their variation and also due to the possibility of speed-accuracy trade-offs. For updating, the correlation between accuracy and *g* was statistically equivalent across updating and non-updating cues. For maintenance, the correlation with *g* for response time on BY trials was significantly stronger on long-delay, relative to short-delay, trials (the *g* correlation was actually negative for the latter type of trial). For task switching, the correlation with *g* for BX trials (on long-delay trials) was significantly stronger on switching trials relative to non-switching trials, but it was also negative (i.e., the opposite direction to what would be expected if task switching predicted the recruitment of *g*). These findings are not robust enough to support the claim that updating, maintenance, or task switching demands are risk factors for the recruitment of *g* in task performance, under the constraints of this task.

Across experiments, the correlational findings show how the recruitment of *g* in task performance increases via increased real-time execution demand—as imposed by response inhibition, updating, and maintenance—in the context of a relatively inefficient task conceptualisation (Experiment 5) but not in the context of a relatively efficient task conceptualisation (Experiment 4). When there is a mixture of inefficient and efficient conceptualisations (Experiment 6), response inhibition and updating do not affect the recruitment of *g*, but maintenance demand does (at least for one measure, i.e., response time on BY trials). In the current experiment, more than half of the sample (58% for the colour shape match task; 56% for the DPX task) selected (and presumably applied to performance) an inefficient colour shape match task model (containing three

or four rules), but the remainder of the sample chose an efficient task model. Thus, although the findings of Experiment 6 do not fully support the idea that task switching is related to Spearman's *g*, a significant involvement of task switching in *g* may be observed in a group that holds only inefficient task conceptualisations (on the grounds that task switching requires reduced activation of one task conceptualisation and increased activation of the other task conceptualisation).

It is also possible that the introduction of a requirement to switch between the performances of two different tasks decreased the likelihood of the involvement of inhibition and maintenance in *g*. Interruption of task performance by a second task may have somehow decreased inhibition demand by reducing the level of prepotency associated with response biases, relative to the build-up of prepotency characteristic of the uninterrupted colour shape match and DPX tasks. The introduction of performance interruptions may have also reduced maintenance load by interfering with, and therefore reducing, the build-up of maintenance load imposed by the long-delay block.

In much the same way as the findings relating to real-time execution demand, some observations supported the task conceptualisation efficiency theory of *g* but other findings did not. Participants were split into high (one-rule or two-rule) and low (three-rule or four-rule) efficiency task conceptualisation groups on the basis of the number of rules contained in their chosen task description. Performance on the experimental tasks did not differ between task conceptualisation efficiency (rule) groups. However, performance on the Culture Fair was significantly poorer in low efficiency, relative to high efficiency, task conceptualisation groups; participants that chose the four-rule colour shape match task conceptualisation had significantly lower *g* than participants that chose the one-, two- or three-rule conceptualisations, and participants that chose the four-rule DPX task conceptualisation had significantly lower *g* than participants that chose the two-rule conceptualisation. This supports the notion that task conceptualisation efficiency is related to *g*. Indeed, moderate and significant

correlations were observed between *g* and the number (one to four) of rules comprised in the selected task conceptualisation. The correlational findings, however, generally provide little support for the importance of task conceptualisation efficiency to the recruitment of *g* in task performance, but issues with sample size exist with these analyses.

These findings may suggest that there is something about the task switching paradigm of Experiment 6, relative to the previous single-task paradigms, that acts to reduce the likelihood of significant rule group differences (and, for that matter, significant *g* correlations to begin with, particularly for critical errors). Alternatively, real relationships and differences between variables may have been masked because group sizes were unequal and some of them were very small (group size ranged from $n = 5$ to $n = 18$) because they were determined by the participants' conceptualisation choice. It is also possible that the organic task representation may not have reflected that which was chosen from the rule sheet (this is indeed one of the strongest criticisms of this approach).

The relative contributions of real-time task execution demand versus task conceptualisation efficiency to the recruitment of *g* in task performance was explored by comparing performance–*g* correlations computed with and without controlling for task conceptualisation efficiency, and comparing conceptualisation–*g* correlations computed with and without controlling for performance measures. For each task, the conceptualisation–*g* correlation (which was significant for both tasks) was unchanged when controlling for each performance measure; this shows that task constraints are not driving the correlation between conceptualisation and *g* in either task, as was found in previous experiments. However, unlike previous experiments, which showed that conceptualisation was driving the correlation between inhibition and *g*, none of the performance–*g* correlations changed in strength significantly when controlling for task

conceptualisation efficiency (the number of rules contained in the selected task conceptualisation).

It is interesting that the only performance measures that were significantly correlated with task conceptualisation efficiency were measures that involved inhibition. Conceptualisation efficiency was significantly correlated with inhibition measures in the colour shape match (critical errors) and DPX (response time to BX trials in long-delay, although this effect fell just short of significance), but not with any other measures. These findings provide further support for the interacting effects of low conceptualisation efficiency and high response inhibition demand (but not other types of real-time execution demand) to the recruitment of *g* in task performance. The reason that performance on AY trials (i.e., the other inhibition measure) was not correlated with task conceptualisation efficiency, which may have been expected on these grounds, is unclear.

In light of the findings of Experiment 6, neither increased real-time task execution demand nor decreased task conceptualisation efficiency was a fundamental risk factor for the recruitment of *g* in task performance. However, the experiment did not warrant complete rejection of the conceptualisation theory of *g*. Task conceptualisation efficiency (which was determined by the participant *prior to* task execution) was not only significantly correlated with *g*, but direct statistical comparisons of Culture Fair error across rule groups revealed that the low efficiency four-rule conceptualisation groups made significantly more Culture Fair errors than the higher efficiency rule groups. Task conceptualisation efficiency was also correlated with performance on measures that involved inhibition (but not other measures), perhaps signalling further the interacting importance of both task conceptualisation efficiency and response inhibition demand to the recruitment of *g* in task performance.

CHAPTER 8

General Discussion

Overview

This closing chapter reviews the findings presented in this body of research in relation to extant literature on the postulated cognitive basis of Spearman's *g*. The chapter commences by reminding the reader of the research questions and the theoretical rationale for addressing them. Second, a summary of the main findings and conclusions drawn from each individual experiment is provided. Third, conceptual conclusions concerning the overarching importance of real-time execution demand (inhibition, maintenance, updating, task switching) versus task modelling demand to the recruitment of *g* are presented alongside their implications. Fourth, the limitations associated with the studies described herein are discussed alongside an agenda for further research in this area. The chapter closes by concluding that task modelling appears more fundamental to the recruitment of *g* than real-time processing or storage factors. However, to hold real theoretical ground, the current findings need to be replicated in studies (a) recruiting a very large cohort of participants from a wide range of populations (e.g., ages, cultures), and (b) assessing the effects of task modelling efficiency in the recruitment of *g* in the context of other types of real-time processing or storage factors using a variety of tasks.

Overview of the research framework and rationale

Chapter 1 introduced the concept of Spearman's general factor (1904, 1927)—a *statistical* construct that explains the finding that people that perform one type of task well tend to perform other types of tasks well—and highlighted how understanding the nature of *g* is fundamental to understanding our ability to operate in complex environments. *g* is, indeed, the biggest source of differences between people as

displayed by its ability to predict a multitude of major life outcomes, such as social mobility and academic attainment (e.g., Colom et al., 2007) and health and survival (Deary et al., 2004). More than 100 years have now passed since the discovery of psychometric *g*, but scientists have yet to reach a consensus regarding its cognitive basis, with some opting to treat *g* as an explanation in itself (e.g., Howe, 1996).

The literature on the cognitive basis of *g* provides an abundance of evidence for a strong relationship between *g* and working memory (e.g., Unsworth et al., 2010). However, the mediating factor in this relationship is still disputed, with evidence existing for the importance of a general executive functioning mechanism (e.g., Ackerman et al., 2002) for particular executive functions such as inhibition (e.g., Unsworth et al., 2009), as well as for storage, instead of processing, factors (e.g., Chuderski et al., 2012). Other evidence challenges a primary role for real-time task demand in the recruitment of *g* in task performance, instead suggesting that correlations between task performance and *g* are dependent on the complexity of the task model, or the mental representation or working memory description of a task (e.g., Duncan et al., 2008). A systematic investigation of the impact of task modelling on the recruitment of *g* in the context of other risk factors for the recruitment of *g*, however, has yet to be undertaken.

To complicate matters further, the source of complexity in a full body of instructions is also unclear. Duncan et al.'s (2008) high-complexity "full" instructions outlined the instructions for two tasks and then encouraged one task to be temporarily discarded, whereas their low-complexity "reduced" instructions group only received instructions for the task to be performed. Full instructions may have therefore increased correlations between task performance with *g* due to the inclusion of (a) an increased number of separately represented requirements (leading, as Duncan et al., 2008, suggest, to some of those requirements being neglected), (b) irrelevant requirements which may cause interference, and/or (c) a requirement to re-model the task as task rules are initially learned.

The primary purpose of this thesis was to systematically explore the *relative* impact of various aspects of working memory demand on the recruitment of *g* in task performance. Two tasks were designed that separated and manipulated the level of demand on two aspects of task modelling (chunking of task requirements into a certain number of rules and re-modelling task requirements) and four aspects of real-time task performance (inhibition of prepotent responses, maintenance of information in working memory across a delay, updating of information in working memory, and switching between two tasks). Both performance, and the strength of correlations between performance and *g* (as measured by scores on the Culture Fair), was compared across high versus low real-time execution demand conditions *in the context of* high versus low task modelling demand conditions. Investigations did not extend to processing speed because the relationship between processing speed and *g* is arguably dependent on other factors (e.g., Jensen, 1998), challenging a robust involvement of processing speed in *g*.

The secondary purpose of this thesis was to identify and define limits in task modelling that effect the recruitment of *g*. The first manipulation of task model complexity involved varying the number of distinct rules (two vs. four) presented in task instructions whilst holding the amount of operative task relevant information, and the task performed, constant. The second manipulation involved varying the inclusion of a requirement to alter the understanding of the task formed on the basis of early specified rules as later rules are presented. These investigations were expected to clarify the source of task modelling complexity that increases the recruitment of *g* in task performance.

Overview of the main findings and preliminary conclusions

The primary aim of Experiment 1 was to explore the impact of response inhibition demand on the recruitment of *g* in task performance, in the context of a two-rule (low-complexity) imposed task model. Participants completed a computer-based task which manipulated the level (present vs. absent) of demand for inhibition of a prepotent but

inappropriate response tendency (across both trials and blocks). Task instructions were presented to participants as two distinct task rules: "respond to items that match in colour *or* shape" and "ignore items that do not match in colour or shape and items that match in both colour *and* shape." Although performance (error) on inhibition conditions (both trial and block) was poorer, relative to non-inhibition conditions, the correlation between performance and *g* (as measured by performance on the Culture Fair test) was, unexpectedly, statistically equivalent. These findings were interpreted as showing that although the presence of response inhibition is a risk factor for increased error, it is not a risk factor for the increased recruitment of *g*, at least when task requirements are represented as two distinct rules.

The main aim of Experiment 2, therefore, was to establish whether the (lack of) relationship between task performance and *g* observed in Experiment 1, would strengthen in the context of a four-rule (high-complexity) imposed task model. Participants completed the same computer-based task, and received the same amount of operative task-relevant information, as the participants in Experiment 1. However, task instructions were presented as four distinct rules: "respond to items that match in colour," "respond to items that match in shape," "ignore items that do not match in colour or shape," and "ignore items that match in both colour *and* shape." As expected, correlations between performance and *g* were strengthened in Experiment 2 relative to Experiment 1, but this finding was limited to measures (both trial and block) involving inhibition. Furthermore, in Experiment 2 only, the correlation with *g* was stronger in the block that contained trials requiring inhibition relative to the block that contained non-inhibition trials only. Performance, however, was statistically equivalent across the two experiments. These findings suggested that the relationship between task performance and *g*, specifically in conditions of response inhibition, depends on the number of represented task rules. However, the relative impact of task modelling versus response inhibition to the recruitment of *g* was not possible to establish from these findings.

Experiment 3 was designed to resolve two issues concerning re-modelling of task requirements. First, it was unclear whether the increase in *g* correlations observed in the four-rule (Experiment 2), relative to the two-rule (Experiment 1), group reflected an increase in the number of distinct rules, or the involvement of an additional requirement to re-model the task[26]. Second, when verbally recalling task rules, some participants in Experiment 2 (generally those with higher *g*) showed evidence for re-modelling of task requirements to form a less complex task model (by chunking task requirements into fewer than four task rules). Participants completed the same computer-based task, and received the same amount of operative task-relevant information and number of rules, as the participants in Experiments 2. However, to remove the requirement to re-model the task, Rules 1 and 2 incorporated the operative word *only*: "respond to items that match *only* in colour" and "respond to items that match *only* in shape" (Rules 3 and 4 were unchanged). Rather than to simply repeat the rules, participants were also encouraged (post execution) to describe task rules as they were represented in their mind; this was assumed to provide a more reliable indication of whether, and which (e.g., those with higher *g*, those that produce less error on the experimental task), participants re-model task rules.

Performance was statistically equivalent across all three experiments (Experiments 1 through 3; with the exception of response time across Experiments 2 and 3), but both increasing the number of rules, and including a requirement to re-model the task, made contributions to the recruitment of *g* in task performance. Regardless of whether task instructions involved re-modelling or not, conditions (trials) requiring inhibition were significantly related to *g* in four-rule, but not two-rule, instruction conditions, and, indeed, correlations between *g* and performance on inhibition conditions (trial and

[26] The process of splitting each of the two rules in Experiment 1 into two rules each to form the four rules in Experiment 2 involved the removal of the operative word *or* from the early rule(s). This effectively removed the early indication that items that matched in either colour or shape (but not both) were important. Thus, when the later rule (rule 2 in two-rule and rule 4 in four-rule) explicitly instructed participants to ignore items that matched in colour *and* shape, the task model formed on the basis of the early rules in four-rule needed to updated to reflect this new, inconsistent, information (whereas the task model formed on the basis of the early rule in two-rule was already consistent with this information).

block) were statistically equivalent across Experiments 2 and 3. However, only in the four-rule group that was additionally required to re-model the task (Experiment 2 but not Experiment 3) were inhibition–*g* correlations *significantly* stronger than (a) inhibition–*g* correlations in Experiment 1, and (b) non-inhibition–*g* correlations (i.e., overall block performance calculated with critical error as a criterion for failure vs. overall block performance in the non-inhibition block). Furthermore, analysis of participants' verbal descriptions of task requirements showed that those participants that stated fewer than four task rules tended to have higher *g*. These results, it was argued, support the view that a large number of task constraints is critical to the recruitment of *g* (in conditions of response inhibition), but that the ability to reconceptualise the task model when learning task rules is also associated with individual differences in *g*.

The primary aim of Experiment 4 was to extend the findings of Experiment 1 by exploring the impact of a number of different types of demand—updating, maintenance, and inhibition of both a prepotent response and a *prepared* prepotent response (both in favour of an alternative response, instead of no response)—on the recruitment of *g* in task performance, in the context of a two-rule (low-complexity) imposed task model. The secondary aim was to examine the relationship between the ability to efficiently re-model task demands (forming a less complex task model) and *g*, using a more objective measure of the number of rules represented in the participants' minds. A second computer-based task was designed which separated and manipulated the level of demand on the four types of real-time demand, and the instructions for this task were presented to participants as two distinct rules: "when the first pattern appears, press red if it is a target or a non-target" and "when the second pattern appears, press green if both patterns were targets but press red if either or both patterns were non-targets." In addition to (subjectively) stating the rules verbally, participants also (objectively) selected a rule format from a sheet containing four rule options comprising one, two, three, or four rules.

Consistent with the findings of Experiment 1, despite the emergence of performance differences between high and low real-time execution demand conditions, very few performance measures in Experiment 4 were significantly correlated with *g*, and manipulations of real-time execution demand did not affect task-sensitivity to *g*. Consistent with the findings of Experiment 3, reconceptualisation of task goals (forming a more efficient task model) remained restricted to people at the higher end of the *g* distribution; furthermore, the number of rules selected was significantly correlated with *g*. These results were interpreted as showing that although high real-time task execution demand (including, but not limited to, response inhibition) is associated with increased error, it is only weakly associated with Spearman's *g* when task requirements are represented as two distinct rules.

The main aim of Experiment 5, therefore, was to extend the findings of Experiment 2 by investigating whether the (low) sensitivity of the DPX task to *g* observed in Experiment 4, would increase in the context of a four-rule (high-complexity) imposed task model. Another important consideration in Experiment 5 was the unresolved issue of the *fundamental* factor in the recruitment of *g* (real-time execution demand vs. task modelling). Participants were presented with four distinct task rules which may have involved a requirement to reconceptualise the task based on early rules as later rules are presented: "when the first pattern appears press red if it is a target," and "when the first pattern appears press red if it is a non-target." Consistent with the findings of Experiment 2, correlations with *g* were stronger (a) when task instructions were presented as four, relative to two, distinct rules, and (b) when real-time execution demand was high relative to low. Another novel finding was that when controlling for real-time performance levels, the correlation between task modelling complexity (the number of rules selected) and *g* remained significant; however, when controlling for task modelling complexity, the correlation between inhibition and *g* was significantly reduced. It was argued that the relationship between task performance and *g* was enhanced by increased real-time performance demand (in the context of a high-

complexity task model), but that complexity of the task model, rather than real-time processing and storage demands per se, was more crucial to the recruitment of *g* in task performance.

Experiment 6 was primarily designed to explore previous investigations under more externally valid conditions. The temporal components of the task model were studied by switching between the two tasks (the colour shape match task of Experiments 1 through 3 and the DPX task of Experiments 4 and 5) at unpredictable intervals. Task model complexity was defined by the number of rules selected by participants (prior to completing the task), instead of encouraged by the format of task instructions. It was proposed that these manipulations were a closer representation to real-life experience in which (a) performance of one task is often interrupted by performance of another, and (b) task models are generally generated internally, rather than explicitly defined by instructions. In contrast to previous studies, correlations between task performance and *g* were not significantly strengthened by increasing task model complexity or increasing the level of real-time execution demand. However, a relationship between task modelling and *g* was suggested by the observations of (a) a significant relationship between the number of rules contained in the rule format selected (from the rule sheet) and accuracy on both the test of *g* and conditions requiring inhibition, and (b) significantly lower *g* scores in participants that selected a rule format comprising more rules (three or four) relative to participants that selected fewer rules (one or two). These findings inferred that task modelling complexity (not real-time task execution demand) was important to the recruitment of *g*[27], under the constraints of this more externally valid task.

Nevertheless, across experiments, consistent patterns of increased performance–*g* correlations in (a) high, relative to low, task model complexity groups, and (b) high task execution demand have been observed in conditions of high-complexity (Experiments

[27] However, due to the nature in which participants were split into task model complexity groups (by actively selecting a rule format containing a certain number [one through four] of rules), there were only a small number of participants in some task model efficiency groups.

2, 3, and 5), but not low- (Experiments 1 and 4) or mixed-complexity (Experiment 6), instructions. Overall, the findings were deemed consistent with a task conceptualisation theory of *g*, with real-time execution demands (particularly inhibition) acting to recruit *g* only when demand on task modelling is high.

Conceptual conclusions and implications

Real-time task execution demand

In the studies presented herein, increasing the level of real-time demand (in the form of inhibition of both a prepotent, and a *prepared* prepotent, response, maintaining a cue across a delay, updating this cue, and switching between the performance of two different tasks) was not always associated with increased *g* correlations. Generally, it was only when task requirements were mentally presented as four task rules (Experiments 2 and 5), and not when requirements were presented as two task rules (Experiments 1 and 4), that conditions associated with high real-time execution demand were (a) significantly correlated with *g*, and (b) *significantly more strongly* correlated with *g* than low real-time execution demand conditions. This latter finding is particularly important because the correlational literature often fails to report the *magnitude* of the increase in correlation co-efficient in one condition relative to another, instead arguing that a greater contribution is reflected by one correlation being numerically bigger than another, or by one correlation reaching significance and another one not.

Chapter 1 highlighted the contradictions in the literature regarding the relationship between *g* and working memory processing demands, particularly inhibition. Some studies report significant correlations between *g* and performance on tasks assumed to tap inhibition (e.g., Polderman et al., 2009; Unsworth et al., 2010), whereas other studies found no such relationship (e.g., Friedman et al., 2006). Similarly, the three main (separate but linked) postulated executive functions (inhibition, updating, and shifting; Miyake et al., 2000) are strongly related to fluid intelligence in some studies

(Salthouse et al., 2003), whereas others have found that only updating (and not inhibition or shifting) is strongly correlated with intelligence measures (Friedman et al., 2006). The findings presented in this thesis provide a candidate explanation for such conflicting evidence; these studies may have differed in terms of how the instructions for the tasks were administered, therefore imposing differing levels of demand on task modelling. Alternatively, they may have been dependent on other factors such as storage (e.g., Colom et al., 2008) or processing speed (e.g., Fry & Hale, 1996; Jensen, 1998; Kail & Salthouse, 1994).

Task modelling demand

This thesis presented four main findings which support a robust relationship between task modelling and Spearman's *g*. The first main finding was that increasing the number of rules presented at task instructions (whilst keeping the amount of operative task-relevant information presented constant) significantly strengthened the recruitment of *g* in task performance, particularly in conditions of high real-time processing demand. Building on the findings of Duncan et al. (2008), the experiments demonstrate a *significant* effect of increasing task model complexity via manipulation of task instructions in strengthening the recruitment of *g* in task performance. Furthermore, these findings show that these effects are limited to conditions in which real-time execution demand (especially inhibition) was high. This is at least consistent with other research demonstrating that the relationship between working memory capacity and storage (speed and accuracy of retrieval from long-term memory) depends on the level of response competition in the task (Conway & Engle, 1994). Indeed, that correlations between task performance and *g* increase significantly in the context of (a) an increase in inhibition demand (only when the task model is complex) and (b) an increase in task modelling complexity (only when inhibition demand is high), suggests some interaction between inhibition and task modelling to the recruitment of *g*. Indeed,

a significant relationship was observed between the number of rules comprised in the selected rule formats and inhibition (as well as *g*).

Published research is consistent with the idea that *g* might reflect the ability to maintain task-relevant information, particularly in interference rich conditions. In a review paper, Conway et al. (2003) concluded that the strong relationship between *g* and working memory capacity reflects the active maintenance of goal-relevant information in the face of salient interference (such as proactive interference, response inhibition, or inhibition of a habitual but inappropriate response). Content analyses of intelligence test items have shown that individual differences in Raven's performance are determined by the ability to manage multiple problem-solving goals in working memory (Carpenter et al. 1990), and the role of working memory capacity in Raven's performance varies as a function of the *level* of saliency of alternative responses (in the answer bank; Jarosz & Wiley, 2010).

The finding that the recruitment of *g* in task performance is dependent on the number of distinct rules that task requirements are embedded in shows that task model complexity can be defined in ways other than presenting additional, unrequired information (Duncan et al., 2008). In contrast to Duncan et al. (2008), here, the amount of operative task-relevant information presented to participants was held constant; what differed across groups was the specific chunking of this information into separately defined rules. Evidence for a relationship between the number of items held in working memory and *g* is provided by Fukuda et al. (2010) who found that the relationship between *g* and working memory capacity (performance on a change detection task) is mediated by the number of items that can be represented in working memory simultaneously (and not by the precision of these representations). Another source of task modelling complexity highlighted by these experiments is the requirement to reconceptulaise task demands in task instructions; this also results in significant *g* correlations, but did not significantly increase the strength of the correlation relative to

the four-rule-ONLY condition.

Indeed, the second main finding supporting the involvement of task modelling in *g* was that participants that reconceptualised the task efficiently (grouped task requirements into a fewer number of rules than explicitly defined by their initial instructions) had significantly higher *g* than participants that did not reconceptualise the task. These same findings were observed when grouping participants into "efficient" and "non-efficient" reconceptualisation groups based on both the number of rules comprised in participants' verbal reports of task requirements and the number of rules contained in participants' rule-format selections. Indeed, the third related but separate piece of evidence in support of a task modelling conceptualisation of *g* is that the number of rules in the participants' rule-format selection was significantly correlated with *g*. Because participants were encouraged to both verbally describe the rules, and select a rule format, based on the way in which the instructions were *represented in their own mind*, these findings provide novel and direct evidence for the theory that *g* may reflect the ability to apply a relatively efficient task model to behaviour.

The fourth, and perhaps the most significant finding, is that the correlation between the number of rules in the participants' rule-format selection and *g* remained when controlling for performance on each measure, suggesting that some aspect of real-time task performance (e.g., the presence of inhibition demand) did not drive this relationship. When viewed against the finding that the inhibition–*g* correlation lost significance and decreased significantly in strength when controlling for the number of rules in the participants' rule-format selection (suggesting that task modelling drove this relationship), this finding highlights the *relative* importance of task modelling over and above inhibition demand to the recruitment of *g*. Indeed, other studies have shown that real-time execution factors—such as attention control (e.g., Chuderski et al., 2012), processing speed, updating, and the control of attention (e.g., Colom et al., 2008)—lose their ability to predict fluid intelligence when short-term storage factors are controlled for. The findings are also consistent with evidence for: (a) the fundamental role of

memory maintenance and retrieval (rather than executive attention or inhibition) in individual differences in working memory capacity (Redick et al., 2011); (b) a strong relationship between *g* and retrieval from secondary memory (such as recalling paired associates; Mogle et al., 2008); (c) the overarching importance of basic and general short-term memory processes (encoding, maintenance, and retrieval) in *g*, even though working memory, updating and short-term memory were all highly related to *g* (Martínez et al., 2011).

The tendency of some participants to efficiently reconceptualise the task predicted better performance on the experimental task (as well as on the Culture Fair). People that reconceptualised the task efficiently (chose a one- or two-rule format) performed better on some measures (irrespective of *g*) than people that did not reconceptualise the task or reconceptualised the task inefficiently (chose a three- or four-rule format). This finding can be explained from a shared but limited resource perspective. If a single "pool" of resources is employed to both process and store information during real-time execution and represent task requirements (e.g., Daneman & Carpenter, 1980; Daneman & Hannon, 2012; Just & Carpenter, 1992), then manipulations of both real-time execution demand and the number of task rules may affect how these resources are used. When the task model is low in complexity, few resources are used for storing task requirements, thus the system has adequate resources to cope with the performance demands of the task, regardless of whether real-time execution demand is high or low. When the task model is high in complexity, fewer resources are available for performance demands because they are employed for representing task requirements, resulting in poorer performance. This is supported by evidence that the ability to inhibit prepotent responses on the antisaccade task (Roberts et al., 1994; Mitchell et al., 2002) and performance on other motor response inhibition tasks (Hester & Garavan, 2005) declines with increasing working memory load.

If the task is re-modelled efficiently (generally only by people with high *g*), and the complexity of the task model is effectively reduced, then more resources are available to

successfully manage the processing and storage demands of the task. Indeed, the four-rule task model (that was encouraged when task requirements were presented as four distinct task rules) may have required reconceptualisation (and therefore recruitment of *g*) in order for the system to have enough resources to cope with real-time task demands, whereas the two-rule task model was already quite concise and required little manipulation. Thus, *g* was significantly more related to performance when the imposed task model was high in complexity (and required reconceptualisation) but not when it was low. When two rules are presented, performance is quite independent of *g* because both high and low participants model the task efficiently (as imposed by task rules) and therefore perform well. However, when four rules are presented, people at the higher end of the *g* distribution perform better than people at the lower end of the *g* distribution (particularly when real-time demand is high) because they are capable of efficiently re-modelling complex task models whereas low *g* people tend to be more "stuck in" to a relatively inflexible (and inefficient) task model. Alternatively, a low-complexity task model, more likely to be formed by individuals with high *g*, may decrease the *perceived* difficulty of the task, which may somehow make actual performance of that task easier, or, possessing higher *g* may result in both efficient task re-modelling and successful task performance.

However, the finding of better performance in four-rule, relative to two-rule, groups was restricted to the DPX task; on the colour shape match task performance was strikingly similar across two- and four-rule, despite elevated recruitment of *g* in four-rule. This finding was consistent across all six experiments (although response time in the non-inhibition block was significantly slower in Experiment 3 relative to Experiment 2). Furthermore, not all DPX measures were associated with better performance in the two-rule group; improvements were restricted to updating cues, non-updating cues, non-inhibition trials (AX and BY) and inhibition (BX) trials in the long-delay block only. It is unclear why a more complex set of instructions acted to increase *g* correlations and decrease performance levels on the DPX task, but only acted to

increase *g* correlations (in the context of unchanged performance levels) on the colour shape match task. Duncan et al. (2008) did find increased goal neglect in their full, relative to simple, instructions condition, and they used a task very similar to the colour shape match task used here. Although these findings suggest that a more efficient task model does not *always* render performance of the task easier, they are not inconsistent with *g* reflecting the ability to efficiently model task rules. The task model may indeed guide task performance, and this capacity may be especially important when tasks are high in complexity (e.g., when demand on response inhibition is present), but a more complex task model may not always result in poorer performance (because some participants can reduce the complexity of the task model).

Perhaps the multiple demand network (Duncan, 2010a), or Dehaene et al.'s (1998) global workspace network, could represent the neurological basis of the task model. Duncan (2010a) suggests that when a novel task is performed, the multiple demand network identifies, separates, and assembles different aspects of the task forming a model of current behavioural goals. He argues that multiple demand activity supports the several separate components involved in fluid intelligence tests (e.g., Carpenter et al., 1990) and the challenge of dividing the task into several solvable sub-parts faced by problem-solving systems (Newell, 1990; Sacerdoti, 1974). Indeed, multiple demand activity depends on the number of varying dimensions involved in a problem (Christoff et al., 2001) and is greater when problems are organised into novel sub-parts of chunks (Bor et al., 2003). Furthermore, in Dumontheil et al. (2010), the presentation of each task rule was associated with activity in parts of the multiple demand network which rapidly returned to baseline in the 10–20 s delay between the presentations of each rule; baseline activity was greater when rules were presented with other rules (full-instructions) relative to when the very same rules were presented at separate points in time (reduced-instructions).

Duncan and colleagues (Duncan et al., 2008; Dumontheil et al., 2010; Duncan et al., 2012; Bhandari & Duncan, 2014) provide much evidence to suggest that task modelling

ability may explain individual differences in *g*. For example, the ability to effectively model task requirements may be capable of explaining the positive manifold observed between cognitive test scores (e.g., Carroll, 1993) and the capacity of *g* to predict a wide range of behaviours (e.g., Spearman, 1904, 1927). Perhaps all complex tasks, irrespective of type, require an efficient task model in order to be completed effectively. For relatively less complex and more automatic tasks, however, efficient task conceptualisation may not be as critical for effective performance. This distinction is indeed reflected in the multiple demand network and Dehaene et al.'s global workspace network in which simple tasks are completed automatically but complex tasks require a complex network of neurons.

Other findings: processing speed

In the present studies, the correlation with *g* and response time was inconsistent and did not adhere to any logical pattern. This supports other research that suggests that the strength of the correlation between response time and *g* depends on other factors, such as practice (Ackerman, 1988), strategy use (Grundick & Kranzler, 2001), increased task complexity by the inclusion of a dual task (Jensen, 1998) or competing task (Fogarty & Stankov, 1995), or by using a battery of, rather than a single, choice reaction time task (Jensen, 1998). Indeed, in a large scale meta-analysis, Sheppard and Vernon (2008) indicated a consistent correlation between mental speed and intelligence of $r = -.24$, supporting a limited role of processing speed in intelligence differences. The controversy surrounding the relationship between processing speed and *g* could also be explained by people using speed-accuracy trade-offs showing that speed is not fundamental to *g* (Deary & Stough, 1996). The claim that speed-accuracy trade-offs are impossible in inspection time tasks, partly because they do not involve a speeded response should, however, be noted (Nettelbeck & Lally, 1976; Deary & Stough, 1996).

Research agenda and limitations

To encourage resolution in how the relationships between *g* and traditional working memory processing and storage factors are affected by task modelling complexity, further investigations should be focussed on controlling and manipulating the complexity of the presentation of task instructions for traditional working memory tasks and traditional measures of memory, attention, and processing speed. Such investigations should not only help to elucidate theory on the cognitive basis of *g*, but should also help to establish any interactions between task modelling and other proposed components of working memory and executive function. However, knowledge of how task modelling affects the recruitment of *g* in the context of different functions is quite redundant unless task modelling is defined in more precise detail. An interesting direction for further investigation is to explore the effects of manipulating other possible sources of complexity in task instructions and how this affects both *g* correlations and performance. The studies presented herein have shown that task model complexity (and correlations with *g*) can be significantly reduced when reducing the number of presented rules and when removing a requirement to re-model the task, but these manipulations were not always associated with better performance. Duncan et al. (2008) and Dumontheil et al. (2010), however, did observe reduced goal neglect (and reduced correlation with *g*) when removing unrequired information from task instructions, which suggests that removing interference from the body of instructions results in easier task performance.

Another possible way to explore task model complexity could be to split task instructions into a set of logically defined rules and a set of illogically defined rules, and explore how this affects performance and *g* correlations. Indeed, an illogically defined representation may require re-modelling (whereas a logical representation may not) which is a requirement that the current findings suggest is important to *g*. It can be argued that the two-rule and four-rule instruction conditions (in both the colour shape match and the DPX tasks) reflected logical and illogical representation of task rules,

respectively. The two-rule instructions for the colour shape task logically defined a go response (Rule 1) and a no-go response (Rule 2). Similarly, the two-rule instructions for the DPX task defined the responses to be made on cues (Rule 1) and the responses to be made on probes (Rule 2). However, splitting each of these rules into two separate rules (to form four-rule instructions), may have effectively split logically-grouped information into more chunks than was necessary. In order to separate out the effect of how logically defined task rules are, further experiments will need to control for the number of rules. For example, the two-rule instructions of the colour shape match task could be made more illogical by grouping the information into two rules that each contain information about a go response and a no-go response. Indeed, understanding methods for modelling tasks so that success is more likely (which may be something which differs between tasks and people) could have a range of implications from everyday life tasks to learning and teaching.

However, although a basic assumption of this research was that two- and four-rule formats reflect efficient and inefficient task conceptualisations, respectively, it is problematic to assume that such linguistic chunking of task rules reflects the actual cognitive representation of task constraints. Duncan et al. (2008) envisage that task models probably contain representations of important stimuli, responses, response timings, trial timings, and stimulus–response rules. A mental model for the DPX task, for example, likely contains representations of the visual characteristics of A cues and X probes (important stimuli), the spatial locations of red and green keys and the speeded nature of responses (important response information), the temporal locations of cue responses and pair responses in a trial (response timings), and the length of short and long delays between cue and probe (trial timings), as well as task requirements (specific stimulus–response rules). Computational modelling and neuroimaging methods are required alongside behavioural data for a clearer understanding of how linguistic rules transform into effective behaviour control programmes.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*(3), 288–318.

Ackerman, P. L., Beier, M. E., & Boyle, M.O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*(4), 567–589.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60.

Alliger, G. M. (1988). Do zero correlations really exist among measures of different intellectual abilities? *Educational and Psychological Measurement, 48*(2), 275–280.

Anderson, M. (1992). *Intelligence and development: A cognitive theory.* Oxford: Blackwell Publishing.

Anderson, B. (1999). Brain size, head size, and intelligence quotient in monozygotic twins: Comment. *Neurology*, *53*(1), 242–244.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49*(A), 5–28.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford: Oxford University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D., & Hitch, G. J. (1974).Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47–90). New York: Academic Press.

Barbey, A. K., Colom, R., & Grafman, J. (2012). Dorsolateral prefrontal contributions to human intelligence. *Neuropsychologia*, *51*(7), 1361–1369.

Barch, D. M., Mitropoulou, V., Harvey, P. D., New, A. S., Silverman, J. M., & Siever, L. J. (2004). Context-processing deficits in schizotypal personality disorder. *Journal of Abnormal Psychology*, *113*(4), 556–568.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100.

Baumeister, A. A. (1998) Intelligence and the "personal equation." *Intelligence*, *26*, 255–265.

Baumeister, A. A., & Kellas, G. (1968). Reaction time and mental retardation. *International Review of Research in Mental Retardation*, *3*, 163–193.

Bethel-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, *8*, 205–238.

Bhandari, A., & Duncan, J. (2014). Goal neglect and knowledge chunking in the construction of novel behaviour. *Cognition*, *130*, 11–30.

Bishop, S. J., Fossella, J., Croucher, C. J., & Duncan, J. (2008). COMT val158met genotype affects recruitment of neural mechanisms supporting fluid intelligence. *Cerebral Cortex*, *18*(9), 2132–2140.

Borkowski, J. G. (1965). Interference effects in short-term memory as a function of level of intelligence. *American Journal of Mental Deficiency*, *70*(3), 458–465.

Brewin, C., & Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behaviour Research and Therapy*, *40*(8), 923–930.

Bright, P. D. (1998). *The control of action: An exploration of Spearman's general factor.* Unpublished doctoral thesis, University of Cambridge, Cambridge, England.

Brown, W., & Thomson, G. H. (1921). *The essentials of mental measurement*. Cambridge: Cambridge University Press.

Browning, P. G., & Gaffan, D. (2008). Prefrontal cortex function in the representation of temporally complex events. *The Journal of Neuroscience*, *28*(15), 3934–3940.

Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'annee Psychologique*, *12*, 191–244.

Binet, A., & Simon, T. (1908). Le developpement de l'intelligence chez les enfants. In R. I. A. Binet & T. S. Simon (Eds.), *The development of intelligence in children* (pp. 182–273). Baltimore: Williams & Wilkins.

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron, 37*(2), 361–367.

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 183–196.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*(6), 547–558.

Carlson, D. (1985). The mesh within a global mesh: a flexible, high-speed organization for parallel computation. *Proceedings of the IEEE International Conference on Supercomputing Systems, Florida*, 618–627.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, *97*(3), 404–431.

Carroll, J. B. (1993). *Human cognitive abilities.* Cambridge: Cambridge University Press.

Carroll, J. B. (1997). Psychmetrics, intelligence, and public perception. *Intelligence, 24*(1), 25–52.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747–749.

Caryl, P. (1994). Early event-related potentials correlate with inspection time and intelligence. *Intelligence*, *18*(1), 15–46.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston, MA: Houghton-Mifflin.

Cattell, R. B., & Cattell, H. E. P. (1973). *Measuring intelligence with the culture fair tests.* Champaign, IL: Institute for Personality and Ability Testing.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage, 14*(5), 1136–1149.

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*, 278–295.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Erlbaum.

Cohen, J., Barch, D., Carter, C., & Servan-Schreiber, D. (1999). Schizophrenic deficits in the processing of context: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology*, *108,* 120–133.

Cohn, S. J., Carlson, J. S., & Jensen, A. R. (1985). Speed of information processing in academically gifted youths. *Personality and Individual Differences, 6*(5), 621–629.

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606.

Colom, R., Escorial, S., Shih, P. C., & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences, 42*(8), 1503–1514.

Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences, 39*(5), 1005–1014.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by '*g*'. *Intelligence, 32*(3), 277–296.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547–552.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*; *123*(4), 354–373.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, *104*, 163–191.

Cowan, N. (1995). *Oxford psychology series: No 26. Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, N. (2005). *Working memory capacity.* Hove, East Sussex: Psychology Press.

Cowan, N., Elliott, E. M., Saults, J.S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100.

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence, 31*(6), 567–587.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences*, *95*(24), 14529–14534.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Darowski, E. S., Helder, E., Zacks, R. T., Hasher, L., & Hambrick, D. Z. (2008). Age-related differences in cognition: The role of distraction control. *Neuropsychology*, *22*(5), 638–644.

Deary, I. J., & Caryl, P. G. (1997). Neuroscience and human intelligence differences. *Trends in* Neurosciences, *20*(8), 365–371.

Deary, I. J., Johnson, W., & Houlihan, L. (2009). Genetic foundations of human intelligence. *Human Genetics*, *126*(1), 215–232.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist*, *51*(6), 599–608.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130–147.

Dempster, F. N. (1991). Inhibitory processes: A neglected dimension of intelligence. *Intelligence*, *15*(2), 157–173.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica*, *101*(2), 395–416.

Dempster, F. N., Corkill, A. J., & Jacobi, K. (1995, November). *Individual Differences in Resistance to Interference.* Paper presented at the Annual Meeting of the Psychonomic Society, Los Angeles.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's advanced progressive matrices: Evidence for multidimensional performance determinants. *Intelligence*, *21*(2), 135–155.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*(6554), 279–281.

Detterman, D. K. (1987). Theoretical notions of intelligence and mental retardation. *American Journal of Mental Deficiency*, 92, 2–11.

Detterman, D. K. (1992). Mopping up: The relation between cognitive abilities and intelligence. *American Journal of Mental Retardation,* 97, 295–301.

Detterman, D. K. (Ed.). (1996). *Current topics in human intelligence: Volume 5. The environment*. Norwood, NJ: Ablex Publishing.

Dickman, S. J., & Meyer, D. E. (1988). Impulsivity and speed–accuracy tradeoffs in information processing. *Journal of Personality and Social Psychology, 54*(2), 274–290.

Dillon, R. F., Pohlmann, J. T., & Lohman, D. F. (1981). A factor analysis of Raven's advanced progressive matrices freed of difficulty factors. *Educational and Psychological Measurement, 41*(4), 1295–1302.

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., . . . Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron, 50*(5), 799–812.

Dumontheil, I., Thompson, R., & Duncan, J. (2010). Assembly and use of new task rules in fronto-parietal cortex. *Journal of Cognitive Neuroscience, 23*(1), 168–182.

Duncan, J. (2010a). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172–179.

Duncan, J. (2010b). *How intelligence happens.* New Haven: Yale University Press.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33*(3), 261–268.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*(3), 257–303.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology*, *14*, 713–741.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., . . . Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General*, *137*(1), 131–148.

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin & Review*, *19*(5), 864–870.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., . . . Emslie, H. (2000). A neural basis for general intelligence. *Science*, *289*(5478), 457–460.

Engle, R.W. (2001). What is working memory capacity? In H.L. Roediger, J.S. Nairne, I. Neath, & A.M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 297–314). Washington, DC: American Psychological Association Press.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 145–199). New York: Academic Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.

Fogarty, G. J., & Stankov, L. (1995). Challenging the "law of diminishing returns." *Intelligence*, *21*(2), 157–174.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(5502), 312–316.

Friedman, N. P., Miyake, A., Corely, R. P., Young, S. E., DeFries. J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679.

Galton, F. (1869). *Hereditary genius: An enquiry into its laws and consequences*. New York: Macmillan.

Gardner, H. (1983). *Frames of Mind.* New York, NY: Basic Books.

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* New York: Basic Books.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the twenty-first century*. New York: Basic Books.

Gevins, A. S., & Cutillo, B. C. (1993). Neuroelectric evidence for distributed processing in human working memory. *Electroencephalography & Clinical Neurophysiology*, *87*, 128–143.

Gimmig, D., Huguet, P., Caverni, J., & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence. *Psychonomic Bulletin & Review*, *13*(6), 1005–1010.

Godefroy, O., Cabaret, M., Petit-Chenal, V., Pruvo, J., & Rousseaux, M. (1999). Control functions of the frontal lobes. Modularity of the central-supervisory system? *Cortex*, *35*(1), 1–20.

Goldman-Rakic, P. S. (1988). Topography of cognition: Parallel distributed networks in primate association cortex. *Annual Review of Neuroscience, 11*(1), 137–156.

Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*, *24*(1), 203–320.

Gray, C.M. (1994). Synchronous oscillations in neuronal systems: of cortical neurons: implications for connectivity, computation and mechanisms and functions. *Journal of Computational Neuroscience*, *1*, 11–38.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322.

Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: Science and ethics. *Nature Reviews Neuroscience*, *5*(6), 471–482.

Grudnik, J. L., & Kranzler, J. H. (2001). Meta-analysis of the relationship between intelligence and inspection time. *Intelligence*, *29*(6), 523–535.

Guilford, J. P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Guthke, J., & Stein, H. (1996). Are learning tests the better version of intelligence tests? *European Journal of Psychological Assessment*, *12*(1), 1–13.

Haier, R. J., Siegel, B., Tang, C., Abel, L., & Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence*, *16*(3), 415–426.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-225). San Diego, CA: Academic Press.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., Pietrini, P. (2001). Distributed and overlapping representation of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2429.

Hertzog, C., & Carter, L. (1988). Sex differences in the structure of intelligence: A confirmatory factor analysis. *Intelligence*, *6*, 287–304.

Hester, R., & Garavan, H. (2005). Working memory and executive function: the influence of content and load on the control of attention. *Memory & Cognition*, *33*(2), 221–233.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Hon, N., Epstein, R. A., Owen, A. M., & Duncan, J. (2006). Frontoparietal activity with minimal decision and control. *The Journal of Neuroscience*, *26*(38), 9805–9809.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, *2*, 41–67.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253–270.

Horn, J. L., Donaldson, G., & Engstrom, R. (1981). Apprehension, memory, and fluid intelligence decline in adulthood. *Research on Aging*, *3*(1), 33–84.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Howe, A. C. (1996). Development of science concepts within a Vygotskian framework. *Science Education*, *80*(1), 35–51.

Hunt, E. (1974). Quote the Raven? Nevermore! In L. Gregg (Ed.), *Knowledge and cognition* (pp. 129-157). Potomac, MD: Erlbaum Associates.

Hunter, J. E. (1983a). *The dimensionality of the General Aptitude Test Battery (GATB) and the dominance of general factors over specific factors in the prediction of job performance for the U.S. Employment Services (USES). Test Research Report No. 44.* Washington, D.C.: US Department of Labor.

Hunter, J. E. (1983b). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In E Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257-266). Hillsdale, New Jersey: Erlbaum.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*, 340–362.

Hunter, I. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, *2*, 447–472.

Jarosz, A. F., & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence*, *40*(5), 427–438.

Jaušovec, N., & Jaušovec, K. (2000). Correlations between ERP parameters and intelligence: A reconsideration. *Biological Psychology*, *55*(2), 137–154.

Jensen, A. R. (1965). Scoring the Stroop test. *Acta Psychologica*, *24*, 398–408.

Jensen, A. R. (1982). The chronometry of intelligence. *Advances in the Psychology of Human Intelligence*, *1*, 255–310.

Jensen, A. R. (1987). Process differences and individual differences in some cognitive tasks. *Intelligence*, *11*(2), 107–136.

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, *13*(8), 869–881.

Jensen, A. R. (1997). Adoption data and two g-related hypotheses. *Intelligence*, *25*, 1–6.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Oxford: Elesvier.

Jensen, A. R., Cohn, S. J., & Cohn, C. M. (1989). Speed of information processing in academically gifted youths and their siblings. *Personality and Individual Differences*, *10*(1), 29–33.

Jensen, A. R., & Rohwer, W. D. (1966). The Stroop color-word test: A review. *Acta Psychologica*, *25*, 36–93.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *14*(89), 81.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2), 199–225.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse, (Eds.), *Variation in working memory* (pp. 21–48). Oxford: Oxford University Press.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217.

Kaufman, A. S., & Horn, J. L. (1996). Age changes on tests of fluid and crystallized ability for women and men on the Kaufman Adolescent and Adult Intelligence Test (KAIT) at ages 17–94 years. *Archives of Clinical Neuropsychology*, *11*(2), 97–121.

Kellman, P. J. (2000). An update on Gestalt psychology. In B. Landau, J. Sabini, E. Newport, & J. Jonides (Eds.), *Essays in honour of Henry and Lila Gleitman* (pp. 157–190). Cambridge, MA: MIT Press.

Kerns, J. G., Cohen, J. D., MacDonald III, A. W., Cho, R. Y., Stenger, A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjestments in control. *Science*, *303*, 1023–1026.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248–294.

Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence,13*(4), 329–347.

Kranzler, J. H., & Jensen, A. R. (1991). Unitary *g*: Unquestioned postulate or empirical fact? *Intelligence*, *15*(4), 437–448.

Kranzler, J. H., Whang, P. A., & Jensen, A. R. (1994). Task complexity and the speed and efficiency of elemental information processing: Another look at the nature of intellectual giftedness. *Contemporary Educational Psychology*, *19*(4), 447–459.

Kusunoki, M., Sigala, N., Gaffan, D., & Duncan, J. (2009). Detection of fixed and variable targets in the monkey prefrontal cortex. *Cerebral Cortex*, *19*(11), 2522–2534.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433.

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*(3), 309–325.

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *The Quarterly Journal of Experimental Psychology: Section A*, *49*(1), 29–50.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527.

Lowe, C., & Rabbitt, P. (1997). Cognitive models of ageing and frontal lobe deficits. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 39–60). Hove, England: Psychology Press.

Luciano, M., Wright, M., Smith, G., Geffen, G., Geffen, L., & Martin, N. (2001). Genetic covariance among measures of information processing speed, working memory, and IQ. *Behavior Genetics*, *31*(6), 581–592.

Luria, A. R. (1966). *Higher cortical functions in man.* London: Tavistock.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199–207.

Lynn, R., Allik, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's standard progressive matrices. *Intelligence*, *32*(4), 411–424.

MacDonald III, A. W., Goghari, V. M., Hicks, B. M., Flory, J. D., Carter, C. S., & Manuck, S. B. (2005). A convergent-divergent approach to context processing, general intellectual functioning, and the genetic liability to schizophrenia. *Neuropsychology*, *19*(6), 814–821.

MacLullich, A., Ferguson, K., Deary, I., Seckl, J., Starr, J., & Wardlaw, J. (2002). Intracranial capacity and brain volumes are associated with cognition in healthy elderly men. *Neurology*, *59*(2), 169–174.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition*, *27*(5), 759–767.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 362–372.

McClure, M. M., Harvey, P. D., Goodman, M., Triebwasser, J., New, A., Koenigsberg, H. W., . . . Siever, L. J. (2010). Pergolide treatment of cognitive deficits associated with schizotypal personality disorder: continued evidence of the importance of the dopamine system in the schizophrenia spectrum. *Neuropsychopharmacology*, *35*(6), 1356–1362.

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136−181). New York: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, *43*(2), 335–354.

Michel, F., & Anderson, M. (2009). Using the antisaccade task to investigate the relationship between the development of inhibition and the development of intelligence. *Developmental Science*, *12*(2), 272–288.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.

Mitchell, J. P., Macrae, C. N., & Gilchrist, I. D. (2002). Working memory and the suppression of reflexive saccades. *Journal of Cognitive Neuroscience*, *14*(1), 95–103.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological Science*, *19*(11), 1071–1077.

Monsell, S. (1996) Control of mental processes. In V. Bruce (Ed.) *Mysteries of the mind: Tutorial essays on cognition* (pp. 93–148). Hove, UK: Erlbaum.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134–140.

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, *31*(3), 327–342.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*, 111–121.

Nettelbeck, T., & Lally, M. (1976). Inspection time and measured intelligence. *British Journal of Psychology*, *67*(1), 17–22.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Norman, D.A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In: Davidson, R.J., Schwartz, G.E., Shapiro, D. (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum Press.

Oberauer, K., Süß, H., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (Eds.), *Variation in working memory* (pp. 49–75). Oxford; New York: Oxford University Press.

Oltmanns, T. F., & Neale, J. M. (1975). Schizophrenic performance when distractors are present: attentional deficit or differential task difficulty? *Journal of Abnormal Psychology*, *84*(3), 205–209.

Pedersen, N. L., Plomin, R., & McClearn, G. E. (1994). Is there G beyond g? (Is there genetic influence on specific cognitive abilities independent of genetic influence on general cognitive ability?). *Intelligence*, *18*(2), 133–143.

Phillips, L. H., & Rabbit, P. (1995). Impulsivity and speed-accuracy strategies in intelligence test performance. *Intelligence, 21*(1), 13–29.

Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science*, *8*(6), 442–447.

Plomin, R., Pedersen, N. L., Lichtenstein, P., & McClearn, G. E. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behavior genetics*, *24*(3), 207–215.

Polderman, T. J., de Geus, E. J., Hoekstra, R. A., Bartels, M., van Leeuwen, M., Verhulst, F. C., . . . Boomsma, D. I. (2009). Attention problems, inhibitory control, and intelligence index overlapping genetic factors: A study in 9-, 12-, and 18-year-old twins. *Neuropsychology*, *23*(3), 381–391.

Rao, S. C., & Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, *276*(5313), 821–824.

Rabbitt, P., & Goward, L. (1994). Age, information processing speed, and intelligence. *The Quarterly Journal of Experimental Psychology*, *47*(3), 741–760.

Raven, J. C. (1965) *Guide to using the Colored Progressive Matrices Sets A, Ab, and B*. Dumfries, Scotland: Griem.

Raven, J. C., Court J. H., Raven, J. (1988). *Manual for Raven's progressive matrices and vocabulary scales.* London: Lewis.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 308–324.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, *1*(3), 86–89.

Reed, T. E., & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence*, *16*(3), 259–272.

Reed, T. E., Vernon, P. A., & Johnson, A. M. (2004). Sex difference in brain nerve conduction velocity in normal humans. *Neuropsychologia*, *42*(12), 1709–1714.

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, *123*(4), 374–393.

Rockstroh, S., & Schweizer, K. (2001). The contributions of memory and attention processes to cognitive abilities. *The Journal of General Psychology*, *128*(1), 30–42.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–321.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*(5), 343–350.

Rushton, J. P., & Ankney, C. D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience*, *119*(5), 692–732.

Rutledge, P., Hollenberg, D., & Hancock, R. A. (1993). Individual differences in the Wegner rebound effect: Evidence for a moderator variable in thought rebound following thought suppression. *Psychological Reports*, *72*(3), 867–880.

Saccuzzo, D. P., Johnson, N. E., & Guertin, T. L. (1994). Information processing in gifted versus nongifted African American, Latino, Filipino, and white children: Speeded versus nonspeeded paradigms. *Intelligence*, *19*(2), 219–243.

Sacerdoti, E. D. (1974). Planning in a hierarchy of abstraction spaces. *Artificial Intelligence*, *5*(2), 115–135.

Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, *5*(5), 479–484.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology*, *84*, 171–199.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*(4), 566–594.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience*, *27*(9), 2349–2356.

Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context: A test of a theoretical model. *Archives of General Psychiatry*, *53*(12), 1105–1112.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, *44*(3), 535–551.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627–670.

Sigala, N., Kusunoki, M., Nimmo-Smith, I., Gaffan, D., & Duncan, J. (2008). Hierarchical coding for sequential task events in the monkey prefrontal cortex. *Proceedings of the National Academy of Sciences, 105*(33), 11969–11974.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268–288.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*(5408), 1657–1661.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–293.

Spearman, C. (1927). *The Abilities of Man.* New York: Macmillan.

Spearman, C. (1938). Measurement of intelligence. *Scientia Milano, 64*, 75–82.

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, *84*(4), 353–378.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J., & Gardner, M. K. (1982). A componential interpretation of the general factor in human intelligence. In Eysenck, H. J. (Ed.), *A model for intelligence* (pp. 231-254). Heidelberg: Springer-Verlag.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Süß, H., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale*. Houghton: Mifflin.

Teuber, H. (1972). University and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis*, *32*, 615–656.

Thompson, R., & Duncan, J. (2009). Attentional modulation of stimulus representation in human fronto-parietal cortex. *NeuroImage*, *48*(2), 436–448.

Thurstone, L. L. (1938). *Primary Mental Abilities.* Chicago: University of Chicago Press.

Thurstone, L. L. (1973). *The nature of intelligence*. Westport, CT: Greenwood Press. (Original work published 1924).

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence, 33*(1), 67–81.

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635–654.

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence, 38*(1), 111–122.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*(4), 392–406.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychology Science Quarterly*, *51*(4), 388–402.

van der Ven, A., & Ellis, J. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Differences, 29*(1), 45–64.

Verguts, T., & Boeck, P. D. (2002). The induction of solution rules in Raven's progressive matrices test. *European Journal of Cognitive Psychology, 14*(4), 521–547.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence*, *7*(1), 53–70.

Vernon, P. A., & Mori, M. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence*, *16*(3), 273–288.

Vernon, P. A., & Weese, S. E. (1993). Predicting intelligence with multiple speed of information-processing tests. *Personality and Individual Differences*, *14*(3), 413–419.

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence*, *34*(3), 261–272.

Walhovd, K. B., & Fjell, A. M. (2007). White matter volume predicts reaction time instability. *Neuropsychologia*, *45*(10), 2277–2284.

Weiskrantz, L. (1997). *Consciousness lost and found: A neuropsychological exploration*. Oxford: Oxford University Press.

Wechsler D. (1997). *Wechsler adult intelligence scale-III*. New York: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-III*. New York: Psychological Corporation.

Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science*, *21*(4), 258–262.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. (2011). New rule use drives the relation between working memory capacity and raven's advanced progressive matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 256–263.

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., . . . Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *107*(33), 14899–14902.

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage*, *56*(2), 744–752.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, *5*, 7–17.

APPENDIX A: Instructions for Experiment 1

Note. *The main text outlines verbatim instructions. Labels for each aspect of the instructions are presented in bold, and actions required by the experimenter are presented in parentheses and upper case. Only the main text was read out loud to the participant with greater verbal emphasis being placed on italicised words.*

Task instructions

Task description. In this task you will see pairs of coloured shapes, one containing a tick and the other containing a cross, coming up one after the other in the middle of the screen. (POINT TO SCREEN AND TAP AT APPROPRIATE RATE.) Here is an example to give you the idea (PLACE EXAMPLE TRIAL IN FRONT OF PARTICIPANT). Imagine these pairs coming up one after the other. As you can see, they can share either colour or shape, or may share neither. The sequence lasts just a few seconds, and near the end, you may also see a pair of items that share both colour and shape (POINT TO 7TH FRAME).

Task rules. In this task, there are two rules you must follow. There are two buttons to press, either left (POINT TO B KEY) or right (POINT TO N KEY). The first rule, the GO rule, is: respond to items that match in colour *or* shape by pressing the side corresponding to the placement of the tick. The second rule, the NOGO rule, is: ignore items that do not match in colour or shape *and* items that match in both colour *and* shape.

Example. We will go through an example in detail now. When I point to a frame, I'd like you to tell me how you would respond. (POINT TO 1ST PAIR) The first pair has a blue square and a red circle – what would you do here? (If 'ignore') Good. (If 'press left' or 'press right') No you would ignore it (REPEAT RULE 2). (POINT TO 2ND PAIR) Both items in the second pair are green – what would you do here? (If 'press right') Good. (If 'ignore' or 'press left') No you would press right (REPEAT RULE 1). (POINT TO 3RD PAIR) What would you do here? (If 'ignore') Good. (If 'press left' or 'press right') No you would ignore it (REPEAT RULE 2). (CONTINUE POINTING TO PAIRS AND CONFIRMING RESPONSES, OR CORRCTING RESPONSES

[AND REPEATING APPROPRIATE RULE IF NECESSARY], FOR REMAINDER OF EXAMPLE.) The whole thing will be fairly quick, so don't be surprised at that.

Learning of task rules

Reminder of task rules. Is that clear? Just to remind you, there are two rules. The first rule is: respond to items that match in colour *or* shape by pressing the key corresponding to the placement of the tick. The second rule is: ignore items that do not match in colour or shape, and items that match in both colour *and* shape.

Participant states rules. Would you repeat the rules please? (If recalled without omissions) That's right (STATE BOTH RULES). (If recalled with omissions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE BOTH RULES).

Task execution

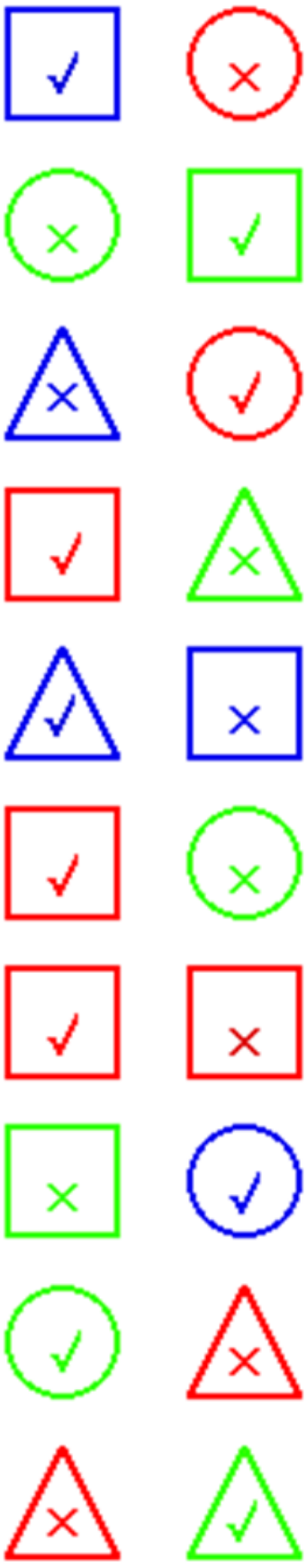
Task execution (block 1). Now remember, this will be very fast – try not to be too surprised and just do your best by responding as quickly and as accurately as you can. Are you ready? (CLICK LEFT MOUSE BUTTON.)

Participant states rules. How many rules were there? Would you repeat the rules please? (If recalled without omissions) That's right (STATE BOTH RULES). (If recalled in a different format to task instructions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE BOTH RULES).

Task execution (block 2). Are you ready to continue? Remember to just do your best by responding as quickly and as accurately as you can. Are you ready? (CLICK LEFT MOUSE BUTTON.)

Participant states rules. How many rules were there? Would you repeat the rules one last time please? (Whether recalled with or without omissions) Okay, thank you.

APPENDIX B: Example Colour Shape Match Trial



APPENDIX C: Instructions for Experiment 2

Note. *The main text outlines verbatim instructions. Labels for each aspect of the instructions are presented in bold, and actions required by the experimenter are presented in parentheses and upper case. Only the main text was read out loud to the participant with greater verbal emphasis being placed on italicised words.*

Task instructions

Task description. In this task you will see pairs of coloured shapes, one containing a tick and the other containing a cross, coming up one after the other in the middle of the screen. (POINT TO SCREEN AND TAP AT APPROPRIATE RATE.) Here is an example to give you the idea (SHOW EXAMPLE). Imagine these pairs coming up one after the other. As you can see, they can share either colour or shape, or may share neither. The sequence lasts just a few seconds, and near the end, you may also see a pair of items that share both colour and shape (POINT TO 7^{TH} FRAME).

Task rules. In this task, there are four rules you must follow. There are two buttons to press, either left (POINT TO KEY B) or right (POINT TO KEY N). The first rule, a GO rule, is: respond to items that match in colour by pressing the side corresponding to the placement of the tick. The second rule, a GO rule, is: respond to items that match in shape by pressing the side corresponding to the placement of the tick. The third rule, a NO GO rule, is: ignore items that do not match in colour or shape. The fourth rule, a NO GO rule, is: ignore items that match in both colour *and* shape.

Example. We will go through an example in detail now. When I point to a frame, I'd like you to tell me how you would respond. (POINT TO 1^{ST} PAIR) The first pair has a blue square and a red circle – what would you do here? (If 'ignore') Good. (If 'press left' or 'press right') No you would ignore it (REPEAT RULE 3). (POINT TO 2^{ND} PAIR) Both items in the second pair are green – what would you do here? (If 'press right') Good. (If 'ignore' or 'press left') No you would press right (REPEAT RULE 1). (POINT TO 3^{RD} PAIR) What would you do here? (If 'ignore') Good. (If 'press left' or 'press right') No you would ignore it (REPEAT RULE 3). (CONTINUE POINTING

TO FRAMES AND CONFIRMING RESPONSES, OR CORRCTING RESPONSES [AND REPEATING APPROPRIATE RULE IF NECESSARY], FOR REMAINDER OF EXAMPLE.) The whole thing will be fairly quick, so don't be surprised at that.

Learning of task rules

Reminder of task rules. Is that clear? Just to remind you, there are four rules. The first rule is: respond to items that match in colour by pressing the side corresponding to the placement of the tick. The second rule is: respond to items that match in shape by pressing the side corresponding to the placement of the tick. The third rule is: ignore items that do not match in colour or shape. The fourth rule is: ignore items that match in both colour *and* shape.

Participant states rules. Would you repeat the rules please? (If recalled without omissions) That's right (STATE ALL FOUR RULES). (If recalled with omissions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE ALL FOUR RULES).

Task execution

Task execution (block 1). Now remember, this will be very fast – try not to be too surprised and just do your best by responding as quickly and as accurately as you can. Are you ready? (CLICK LEFT MOUSE BUTTON.)

Participant states rules. How many rules were there? Would you repeat the rules please? (If recalled without omissions) That's right (STATE ALL FOUR RULES). (If recalled in a different format to task instructions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE ALL FOUR RULES).

Task execution (block 2). Are you ready to continue? Remember to just do your best by responding as quickly and as accurately as you can. Are you ready? (CLICK LEFT MOUSE BUTTON.)

Participant states rules. How many rules were there? Would you repeat the rules one last time please? (Whether recalled with or without omissions) Okay, thank you.

APPENDIX D: Instructions for Experiment 3

Note. *The main text outlines verbatim instructions. Labels for each aspect of the instructions are presented in bold, and actions required by the experimenter are presented in parentheses and upper case. Only the main text was read out loud to the participant with greater verbal emphasis being placed on italicised words.*

Task instructions

Task description. In this task you will see pairs of coloured shapes, one containing a tick and the other containing a cross, coming up one after the other in the middle of the screen. (POINT TO SCREEN AND TAP AT APPROPRIATE RATE.) Here is an example to give you the idea (SHOW EXAMPLE). Imagine these pairs coming up one after the other. As you can see, they can share either colour or shape, or may share neither. The sequence lasts just a few seconds, and near the end, you may also see a pair of items that share both colour and shape (POINT TO 7^{TH} FRAME).

Task rules. In this task, there are four rules you must follow. There are two buttons to press, either left (POINT TO KEY B) or right (POINT TO KEY N). The first rule, a GO rule, is: respond to items that match *only* in colour by pressing the side corresponding to the placement of the tick. The second rule, a GO rule, is: respond to items that match *only* in shape by pressing the side corresponding to the placement of the tick. The third rule, a NO GO rule, is: ignore items that do not match in colour or shape. The fourth rule, a NO GO rule, is: ignore items that match in both colour *and* shape.

Example. We will go through an example in detail now. When I point to a frame, I'd like you to tell me how you would respond. (POINT TO 1^{ST} PAIR) The first pair has a blue square and a red circle – what would you do here? (If 'ignore') Good. (If 'press left' or 'press right') No you would ignore it (REPEAT RULE 3). (POINT TO 2^{ND} PAIR) Both items in the second pair are green – what would you do here? (If 'press right') Good. (If 'ignore' or 'press left') No you would press right (REPEAT RULE 1). (POINT TO 3^{RD} PAIR) What would you do here? (If 'ignore') Good. (If 'press left' or

'press right') No you would ignore it (REPEAT RULE 3). (CONTINUE POINTING TO FRAMES AND CONFIRMING RESPONSES, OR CORRCTING RESPONSES [AND REPEATING APPROPRIATE RULE IF NECESSARY], FOR REMAINDER OF EXAMPLE.) The whole thing will be fairly quick, so don't be surprised at that.

Learning of task rules

Reminder of task rules. Is that clear? Just to remind you, there are four rules. The first rule is: respond to items that match *only* in colour by pressing the side corresponding to the placement of the tick. The second rule is: respond to items that match *only* in shape by pressing the side corresponding to the placement of the tick. The third rule is: ignore items that do not match in colour or shape. The fourth rule is: ignore items that match in both colour *and* shape.

Participant states rules. Would you repeat the rules please? (If recalled without omissions) That's right (STATE ALL FOUR RULES). (If recalled with omissions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE ALL FOUR RULES).

Task execution

Task execution (block 1). Now remember, this will be very fast – try not to be too surprised and just do your best by responding as quickly and as accurately as you can. Are you ready? (CLICK LEFT MOUSE BUTTON.)

Participant states rules. How many rules were there? Would you repeat the rules please? (If recalled without omissions) That's right (STATE ALL FOUR RULES). If recalled in a different format to task instructions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE ALL FOUR RULES).

Task execution (block 2). Are you ready to continue? Remember to just do your best by responding as quickly and as accurately as you can. Are you ready? (CLICK LEFT MOUSE BUTTON.)

Reconceptualisation of task rules

Let's think about the rules of the task one more time. I'd like you to repeat the rules again, but this time, I'd like you to describe the rules in the way that they are represented in your *own mind*, which may be the same as, or different to, my description of the rules.

So I described the task as having four rules. (STATE ALL FOUR RULES).

Try to imagine that you are the experimenter and I am the participant who is about to carry out the task. I'd like you to say the rules in the way that best reflects how the rules are represented in your *own mind*. How would you present the rules and how many rules would there be? Please take as long as you need to think carefully about the way in which the instructions are represented *in your own mind.*

APPENDIX E: Rule Sheet for Experiments 4 and 5

Tick the set of rules that most reflects *how <u>you thought</u> about the task, or write your own description*

☐	In my head, the rule information is in one part. The rule is that when the first pattern appears press red if it is a TARGET or a NON-TARGET and when the second pattern appears press green if BOTH patterns were TARGETS and press red if EITHER or BOTH patterns were NON-TARGETS.
☐	In my head, the rule information is in two parts. The first rule is that when the first pattern appears press red if it is a TARGET or a NON-TARGET. The second rule is that when the second pattern appears press green if BOTH patterns were TARGETS and press red if EITHER or BOTH patterns were NON-TARGETS.
☐	In my head, the rule information is in three parts. The first rule is that when the first pattern appears press red if it is a TARGET or a NON-TARGET. The second rule is that when the second pattern appears press green if BOTH patterns were TARGETS. The third rule is that when the second pattern appears press red if EITHER or BOTH patterns were NON-TARGETS.
☐	In my head, the rule information is in four parts. The first rule is that when the first pattern appears press red if it is a TARGET. The second rule is that when the first pattern appears press red if it is a NON-TARGET. The third rule is that when the second pattern appears press green if BOTH patterns were TARGETS. The forth rule is that when the second pattern appears press red if EITHER or BOTH patterns were NON-TARGETS.
☐	In my head, the rule information is as follows… _____ _____ _____ _____

APPENDIX F: Instructions for Experiment 4

Note. *The main text outlines verbatim instructions. Labels for each aspect of the instructions are presented in bold, and actions required by the experimenter are presented in parentheses and upper case. Only the main text was read out loud to the participant with greater verbal emphasis being placed on italicised words.*

Task instructions

Task description. In this task you will see a series of dot patterns coming up, one after the other, in the middle of the screen. Here is an example to give you the idea (PLACE EXAMPLE RUN IN FRONT OF PARTICIPANT). You need to view these patterns in pairs. So, (POINT TO 1ST FRAME) a fixation cross will appear indicating the start of the pair, (POINT TO 2ND FRAME) then the first pattern of the pair will appear, (POINT TO 3RD FRAME) then there will be a delay in which the screen is blank, (POINT TO 4TH FRAME) then the second pattern of the pair will appear, (POINT TO 5TH FRAME) then there will be another blank-screen gap before the next fixation cross appears indicating the start of the next pair.

(PLACE EXAMPLE PAIRS IN FRONT OF THE PARTICIPANT) Your task is to look out for the target pair, (POINT TO TARGET PAIR) which is this combination of patterns here. So whenever this pattern (POINT TO 'TARGET FIRST' IN TARGET PAIR) is followed by this pattern (POINT TO 'TARGET SECOND' IN TARGET PAIR) it indicates a target pair. We will call this pattern 'target first' (POINT TO 'TARGET FIRST' IN TARGET PAIR) and this pattern 'target second' (POINT TO TARGET SECOND). All other combinations of patterns are non-target pairs. Notice that non-target pairs could have something in common with the target pair: they could have a 'target first' (POINT TO TARGET FIRST IN TARGET PAIR AND THEN TARGET FIRST IN 1ST NON-TARGET PAIR) but a 'non-target second' (POINT TO NON-TARGET SECOND IN 1ST NON-TARGET PAIR); they could have a 'non-target first' (POINT TO NON-TARGET FIRST IN 2ND NON-TARGET PAIR) but a 'target second' (POINT TO TARGET SECOND IN TARGET PAIR AND THEN TARGET

SECOND IN 2ND NON-TARGET PAIR); or they could have two non-targets (POINT TO NON-TARGET FIRST THEN NON-TARGET SECOND IN 3RD NON-TARGET PAIR).

Task rules. In this task, there are two rules you must follow. There are two buttons to press, either red (POINT TO RED KEY [N]) or green (POINT TO GREEN KEY [B]). The first rule is: respond to the first pattern of the pair by pressing the red key if it is a target or a non-target. The second rule is: respond to the second pattern of the pair by pressing the green key if both patterns were targets and by pressing the red key if either or both patterns were non-targets.

Example. We will go through an example in detail now (ENSURE EXAMPLE RUN IS IN FRONT OF PARTICIPANT). When I point to a frame, I'd like you to tell me how you would respond. First you see a fixation cross indicating the start of the pair. (POINT TO 1ST PATTERN) Then you see the first pattern of the pair – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 1). Then there is a blank-screen delay. (POINT 2ND PATTERN) Then you will see the second pattern of the pair – what would you do here? (If 'press green') Good. (If 'press red') No you would press green (REPEAT RULE 2). Then there is a blank-screen gap before the next pair. In this next pair there is a fixation cross, (POINT TO 3RD PATTERN) then the first pattern – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 1). Then there is a delay, (POINT TO 4TH PATTERN) then the second pattern – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 2). (CONTINUE POINTING TO PATTERNS AND CONFIRMING RESPONSES, OR CORRCTING RESPONSES [AND REPEATING APPROPRIATE RULE IF NECESSARY], FOR REMAINDER OF EXAMPLE.)

Learning of task rules

Reminder of task rules. Is that clear? Just to remind you, there are two rules. The first rule is: respond to the first pattern of the pair by pressing the red key if it is a target or a non-target. The second rule is: respond to the second pattern of the pair by pressing

the green key if both patterns were targets and by pressing the red key if either or both patterns were non-targets.

Participant states rules. Would you repeat the rules please? (If recalled without omissions) That's right (STATE BOTH RULES). (If recalled with omissions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE BOTH RULES).

Task execution

Task execution (block 1). Some of this will be very fast – try not to be too surprised and just do your best by responding as quickly and as accurately as you can. Press the space bar to start.

Participant states rules. Would you repeat the rules please? (If recalled without omissions) That's right (STATE BOTH RULES). (If recalled with omissions or in a different format to that outlined in instructions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE BOTH RULES).

Task execution (block 2). Are you ready to continue? Remember to just do your best by responding as quickly and as accurately as you can. Press the space bar to start.

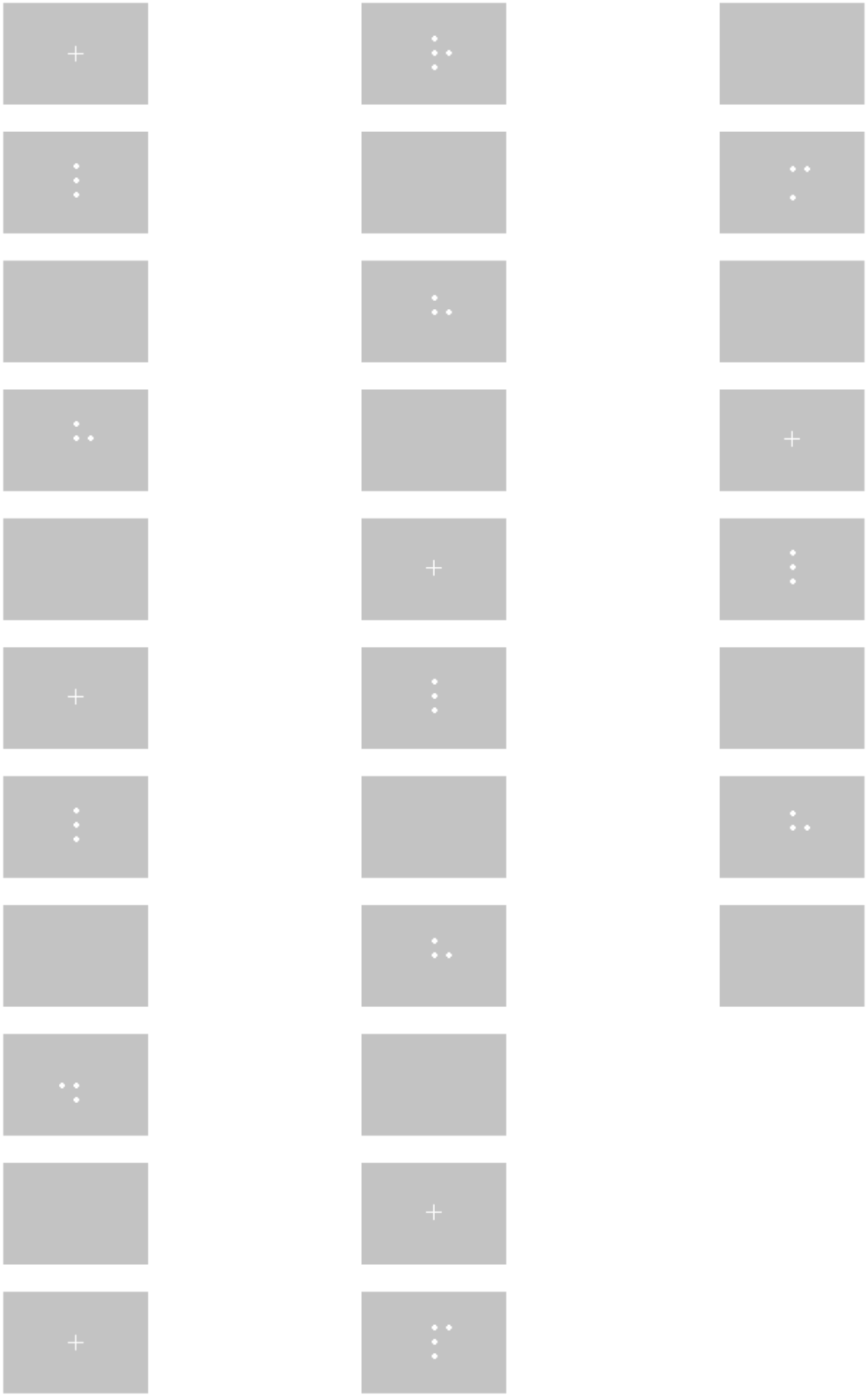
Reconceptualisation of task rules

Let's think about the rules of the task one more time. This time, instead of repeating the rules out loud, I'd like you to think about the way in which the rules are represented in your *own mind*, which may be the same as, or different to, my description of the rules.

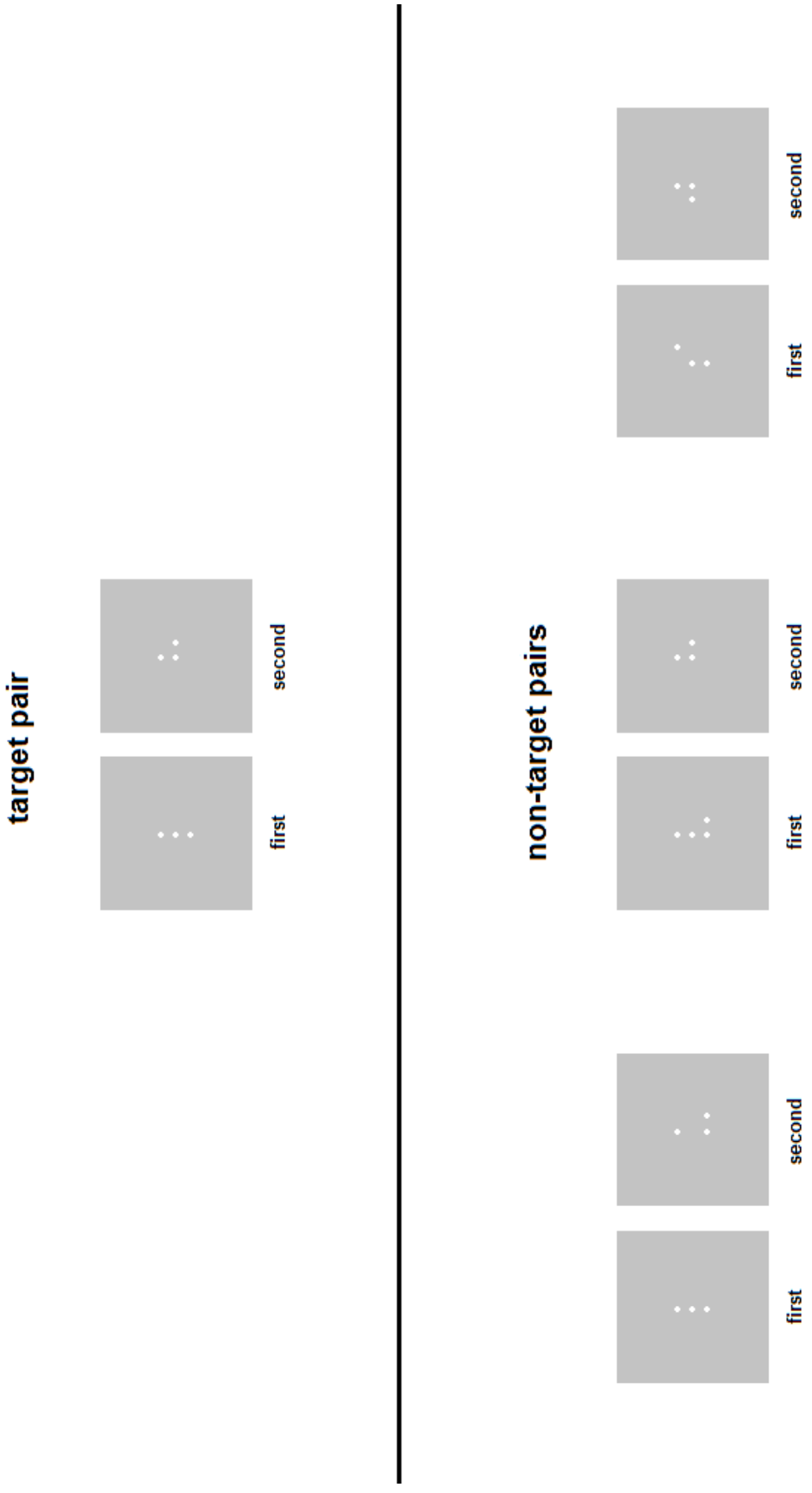
So I described the task as having two rules. (STATE BOTH RULES.)

Here is a short list of possible instruction rules for the task (PLACE RULE SHEET IN FRONT OF PARTICIPANT). I'd like you to read this list and select the option that best reflects how the rules represented in your *own mind*. If none of the options are appropriate you can write your own description and suggest how many distinct rules there are at the bottom. Please take as long as you need to think carefully about the way in which the instructions are represented *in your own mind*.

APPENDIX G: Example DPX Trials for Experiment 4



APPENDIX H: Example DPX Pairs for Experiments 4 and 5



APPENDIX I: Instructions for Experiment 5

Note. *The main text outlines verbatim instructions. Labels for each aspect of the instructions are presented in bold, and actions required by the experimenter are presented in parentheses and upper case. Only the main text was read out loud to the participant with greater verbal emphasis being placed on italicised words.*

Task instructions

Task description. In this task you will see a series of dot patterns coming up, one after the other, in the middle of the screen. Here is an example to give you the idea (PLACE EXAMPLE RUN IN FRONT OF PARTICIPANT). You need to view these patterns in pairs. So, (POINT TO 1ST FRAME) a fixation cross will appear indicating the start of the pair, (POINT TO 2ND FRAME) then the first pattern of the pair will appear, (POINT TO 3RD FRAME) then there will be a delay in which the screen is blank, (POINT TO 4TH FRAME) then the second pattern of the pair will appear, (POINT TO 5TH FRAME) then there will be another blank-screen gap before the next fixation cross appears indicating the start of the next pair.

(PLACE EXAMPLE PAIRS IN FRONT OF THE PARTICIPANT) Your task is to look out for the target pair, (POINT TO TARGET PAIR) which is this combination of patterns here. So whenever this pattern (POINT TO 'TARGET FIRST' IN TARGET PAIR) is followed by this pattern (POINT TO 'TARGET SECOND' IN TARGET PAIR) it indicates a target pair. We will call this pattern 'target first' (POINT TO 'TARGET FIRST' IN TARGET PAIR) and this pattern 'target second' (POINT TO TARGET SECOND). All other combinations of patterns are non-target pairs. Notice that non-target pairs could have something in common with the target pair: they could have a 'target first' (POINT TO TARGET FIRST IN TARGET PAIR AND THEN TARGET FIRST IN 1ST NON-TARGET PAIR) but a 'non-target second' (POINT TO NON-TARGET SECOND IN 1ST NON-TARGET PAIR); they could have a 'non-target first' (POINT TO NON-TARGET FIRST IN 2ND NON-TARGET PAIR) but a 'target second' (POINT TO TARGET SECOND IN TARGET PAIR AND THEN TARGET

SECOND IN 2ND NON-TARGET PAIR); or they could have two non-targets (POINT TO NON-TARGET FIRST THEN NON-TARGET SECOND IN 3RD NON-TARGET PAIR).

Task rules. In this task, there are four rules you must follow. There are two buttons to press, either red (POINT TO RED KEY [N]) or green (POINT TO GREEN KEY [B]). The first rule is: respond to the first pattern of the pair by pressing the red key if it is a target. The second rule is: respond to the first pattern by pressing the red key if it is a non-target. The third rule is: respond to the second pattern of the pair by pressing the green key if both patterns were targets. The fourth rule is: respond to the second pattern by pressing the red key if either or both patterns were non-targets.

Example. We will go through an example in detail now (ENSURE EXAMPLE RUN IS IN FRONT OF PARTICIPANT). When I point to a frame, I'd like you to tell me how you would respond. First you see a fixation cross indicating the start of the pair. (POINT TO 1ST PATTERN) Then you see the first pattern of the pair – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 1). Then there is a blank-screen delay. (POINT 2ND PATTERN) Then you will see the second pattern of the pair – what would you do here? (If 'press green') Good. (If 'press red') No you would press green (REPEAT RULE 3). Then there is a blank-screen gap before the next pair. In this next pair there is a fixation cross, (POINT TO 3RD PATTERN) then the first pattern – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 1). Then there is a delay, (POINT TO 4TH PATTERN) then the second pattern – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 4). (CONTINUE POINTING TO PATTERNS AND CONFIRMING RESPONSES, OR CORRCTING RESPONSES [AND REPEATING APPROPRIATE RULE IF NECESSARY], FOR REMAINDER OF EXAMPLE.)

Learning of task rules

Reminder of task rules. Is that clear? Just to remind you, there are four rules. The first rule is: respond to the first pattern of the pair by pressing the red key if it is a target. The second rule is: respond to the first pattern of the pair by pressing the red key if it is a target. The third rule is: respond to the second pattern of the pair by pressing the green if both patterns were targets. The fourth rule is: respond to the second pattern of the pair by pressing the red key if either or both patterns were non-targets.

Participant repeats rules. Would you repeat the rules please? (If recalled without omissions) That's right (STATE ALL FOUR RULES). (If recalled with omissions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE ALL FOUR RULES).

Task execution

Task execution (block 1). Some of this will be very fast – try not to be too surprised and just do your best by responding as quickly and as accurately as you can. Press the space bar to start.

Participant states rules. Would you repeat the rules please? (If recalled without omissions) That's right (STATE ALL FOUR RULES). (If recalled with omissions or in a different format to that outlined in instructions) That's not quite right (STATE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE ALL FOUR RULES).

Task execution (block 2). Are you ready to continue? Remember to just do your best by responding as quickly and as accurately as you can. Press the space bar to start.

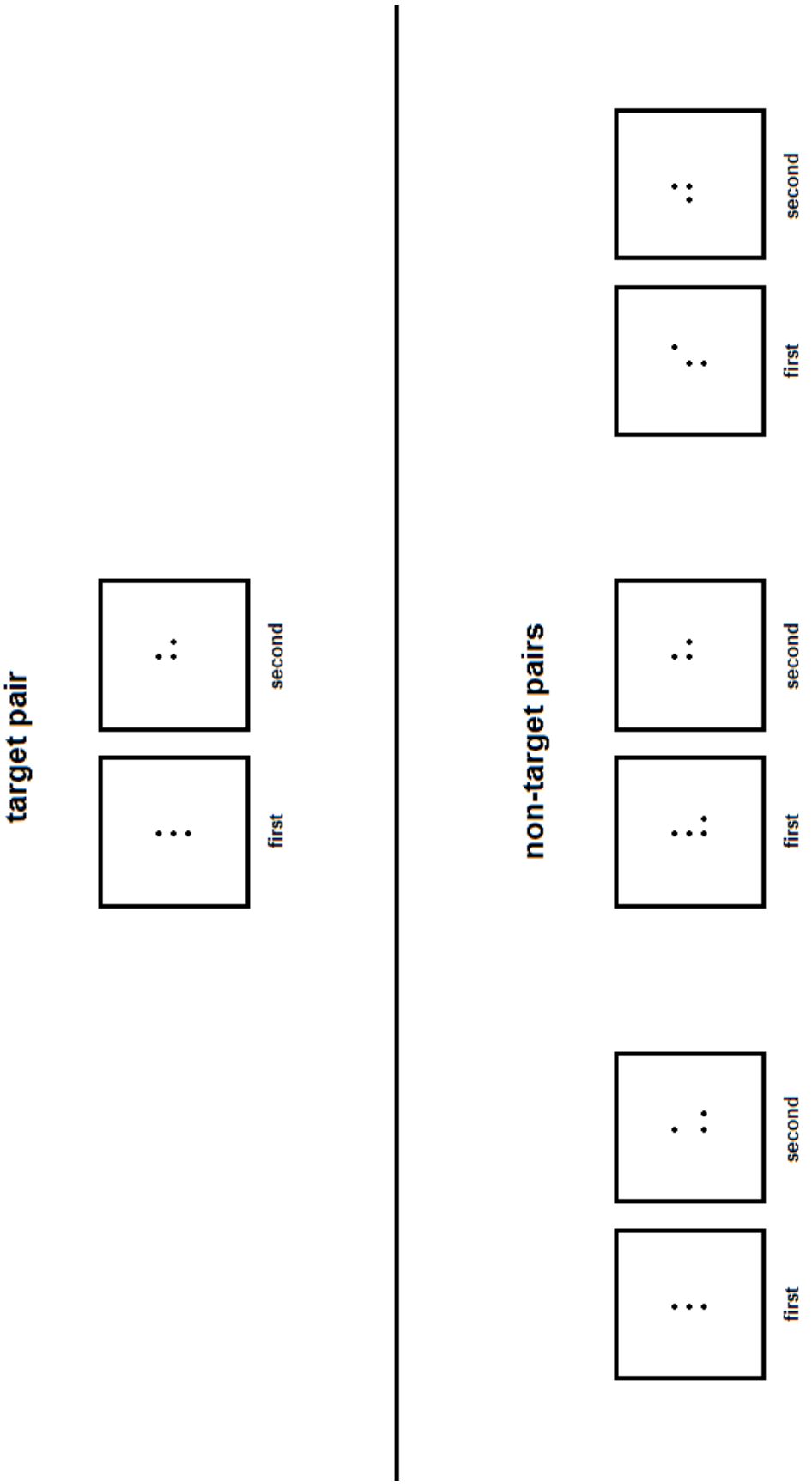
Reconceptualisation of task rules

Let's think about the rules of the task one more time. This time, instead of repeating the rules out loud, I'd like you to think about the way in which the rules are represented in your *own mind*, which may be the same as, or different to, my description of the rules.

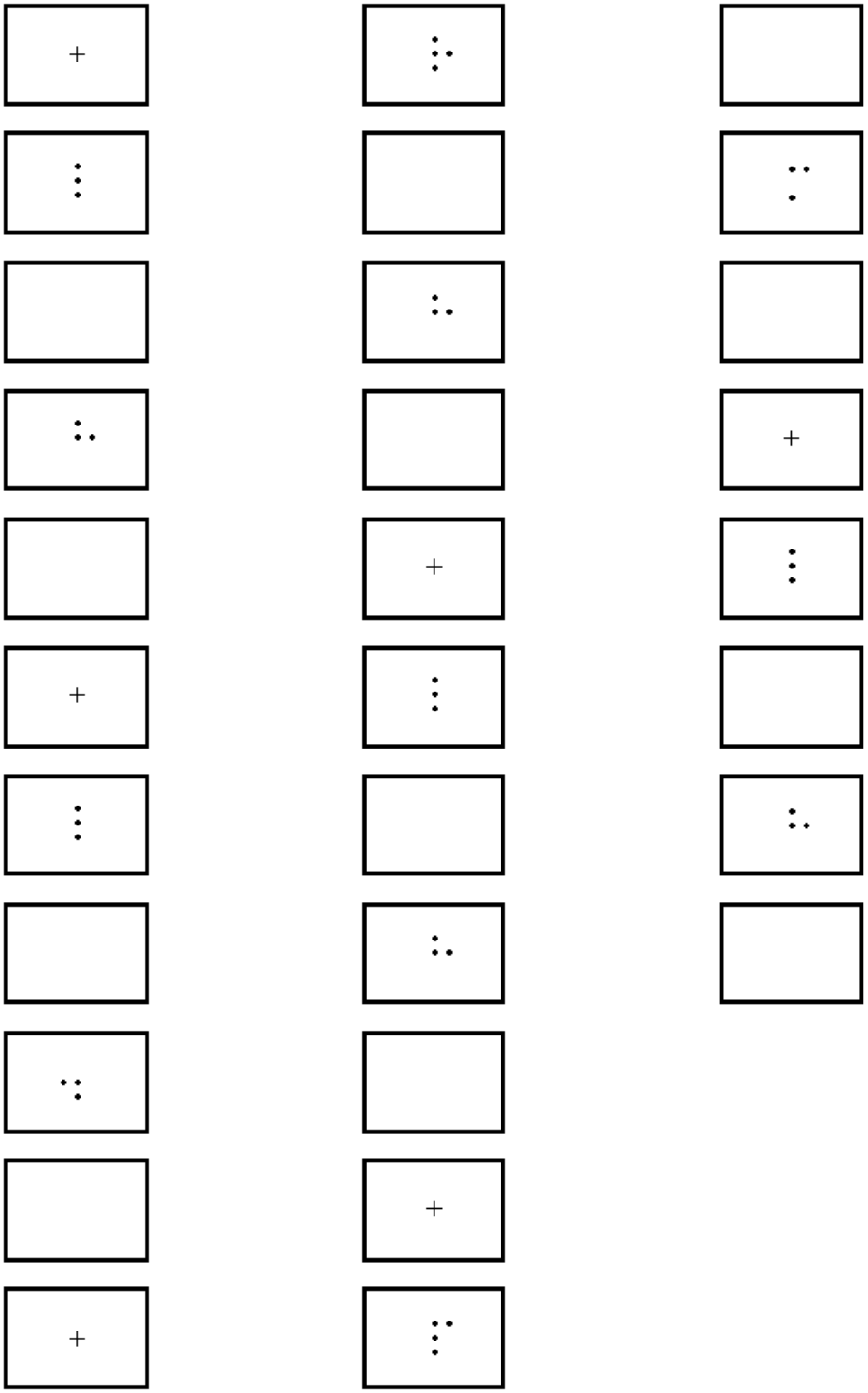
So I described the task as having four rules. (STATE ALL FOUR RULES.)

Here is a short list of possible instruction rules for the task (PLACE RULE SHEET IN FRONT OF PARTICIPANT). I'd like you to read this list and select the option that best reflects how the rules represented in your *own mind*. If none of the options are appropriate you can write your own description and suggest how many distinct rules there are at the bottom. Please take as long as you need to think carefully about the way in which the instructions are represented *in your own mind*.

APPENDIX J: Example DPX Pairs for Experiment 6



APPENDIX K: Example DPX Trials for Experiment 6



APPENDIX L: Colour Shape Match Rule Sheet for Experiment 6

Tick the set of rules that most reflects *how <u>you thought</u> about the task, or write your own description*

☐ **In my head, task instructions are in four parts:**

Part 1: Respond to items that match in colour by pressing the side of the tick

Part 2: Respond to items that match in shape by pressing the side of the tick

Part 3: Ignore items that do not match in colour or shape

Part 4: Ignore items that match in both colour and shape

☐ **In my head, task instructions are in three parts:**

Part 1: Respond to items that match in colour by pressing the side of the tick

Part 2: Respond to items that match in shape by pressing the side of the tick

Part 3: Ignore items that do not match in colour or shape and items that match in both colour and shape

☐ **In my head, task instructions are in three parts:**

Part 1: Respond to items that match in colour or in shape by pressing the side of the tick

Part 2: Ignore items that do not match in colour or shape

Part 3: Ignore items that match in both colour and shape

☐ **In my head, task instructions are in two parts:**

Part 1: Respond to items that match in colour or in shape by pressing the side of the tick

Part 2: Ignore items that do not match in colour or shape and items that match in both colour and shape

☐ **In my head, task instructions are in one part:**

Respond to items that match in colour or in shape by pressing the side of the tick and ignore items that do not match in colour or shape and items that match in both colour and shape

☐ **None of the above. In my head, the rule information is as follows:**

_____

_____

_____

_____

APPENDIX M: DPX Rule Sheet for Experiment 6

Tick the set of rules that most reflects *how <u>you thought</u> about the task, or write your own description*

☐ **In my head, task instructions are in four parts:**

Part 1: Respond to the first pattern by pressing the red key if it is a target

Part 2: Respond to the first pattern by pressing the red key if it is a non-target

Part 3: Respond to the second pattern by pressing the green key if both patterns were targets

Part 4: Respond to the second pattern by pressing the red key if either or both patterns were non-targets

☐ **In my head, task instructions are in three parts:**

Part 1: Respond to the first pattern by pressing the red key if it is a target or a non-target

Part 2: Respond to the second pattern by pressing the green key if both patterns were targets

Part 3: Respond to the second pattern by pressing the red key if either or both patterns were non-targets

☐ **In my head, task instructions are in two parts:**

Part 1: Respond to the first pattern by pressing the red key if it is a target or a non-target

Part 2: Respond to the second pattern by pressing the green key if both patterns were targets and by pressing the red key if either or both patterns were non-targets

☐ **In my head, task instructions are in one part:**

Respond to the first pattern by pressing the red key if it is a target or a non-target and respond to the second pattern by pressing the green key if both patterns were targets and by pressing the red key if either or both patterns were non-targets

☐ **None of the above. In my head, the rule information is as follows:**

_____

_____

_____

_____

APPENDIX N: Instructions for Experiment 6

Note. *The main text outlines verbatim instructions. Labels for each aspect of the instructions are presented in bold, and actions required by the experimenter are presented in parentheses and upper case. Only the main text was read out loud to the participant with greater verbal emphasis being placed on italicised words.*

Introduction

This task contains a mixture of two different tasks. In one of the tasks you will respond to pairs of (if the first block executed contains colour shape match trials) coloured shape/ (if the first block executed contains DPX trials) dot patterns. In the other task you will respond to pairs of (if the first block executed contains colour shape match trials) dot patterns/ (if the first block executed contains DPX trials) coloured shapes. For each task you will make a response by pressing one of these two keys (POINT TO KEYS B AND N), but the specific response will differ depending on which task it is. I'll give you the instructions for each task in turn to ensure that you know exactly what to do.

Colour shape match task

(PRESENT FIRST IF THE FIRST EXECUTED TRIAL IS COLOUR SHAPE MATCH.)

Colour shape match task description. In this task you will see pairs of coloured shapes, one containing a tick and the other containing a cross, coming up one after the other in the middle of the screen. (POINT TO SCREEN AND TAP AT APPROPRIATE RATE.) Here is an example to give you the idea (SHOW EXAMPLE). Imagine these pairs coming up one after the other. As you can see, they can share either colour or shape, or may share neither. The sequence lasts just a few seconds, and near the end, you may also see a pair of items that share both colour and shape (POINT TO 7TH FRAME).

Colour shape match task rules. In this task, there are four rules you must follow. There are two buttons to press, either left (POINT TO KEY B) or right (POINT TO KEY N). The first rule, a GO rule, is: respond to items that match in colour by pressing

the side corresponding to the placement of the tick. The second rule, a GO rule, is: respond to items that match in shape by pressing the side corresponding to the placement of the tick. The third rule, a NO GO rule, is: ignore items that do not match in colour or shape. The fourth rule, a NO GO rule, is: ignore items that match in both colour *and* shape.

Colour shape match example. We will go through an example in detail now. When I point to a frame, I'd like you to tell me how you would respond. (POINT TO 1ST PAIR) The first pair has a blue square and a red circle – what would you do here? (If 'ignore') Good. (If 'press left' or 'press right') No you would ignore it (REPEAT RULE 3). (POINT TO 2ND PAIR) Both items in the second pair are green – what would you do here? (If 'press right') Good. (If 'ignore' or 'press left') No you would press right (REPEAT RULE 1). (POINT TO 3RD PAIR) What would you do here? (If 'ignore') Good. (If 'press left' or 'press right') No you would ignore it (REPEAT RULE 3). (CONTINUE POINTING TO FRAMES AND CONFIRMING RESPONSES, OR CORRCTING RESPONSES [AND REPEATING APPROPRIATE RULE IF NECESSARY], FOR REMAINDER OF EXAMPLE.) The whole thing will be fairly quick, so don't be surprised at that.

Reconceptualisation of colour shape match task rules. I would like you to now think about the way in which the rules are represented in your *own mind*, which may be the same as, or different to, my description of the rules.

So I described the task as having four rules. (STATE ALL FOUR RULES.)

Here is a short list of possible instruction rules for the task (PLACE RULE SHEET IN FRONT OF PARTICIPANT). I'd like you to read this list and select the option that best reflects how the rules represented in your *own mind*. If none of the options are appropriate you can write your own description and suggest how many distinct rules there are at the bottom. Please take as long as you need to think carefully about the way in which the instructions are represented *in your own mind*.

DPX task

(PRESENT FIRST IF THE FIRST EXECUTED TRIALS IS DPX.)

DPX task description. In this task you will see a series of dot patterns coming up, one after the other, in the middle of the screen. Here is an example to give you the idea (PLACE EXAMPLE RUN IN FRONT OF PARTICIPANT). You need to view these patterns in pairs. So, (POINT TO 1ST FRAME) a fixation cross will appear indicating the start of the pair, (POINT TO 2ND FRAME) then the first pattern of the pair will appear, (POINT TO 3RD FRAME) then there will be a delay in which the screen is blank, (POINT TO 4TH FRAME) then the second pattern of the pair will appear, (POINT TO 5TH FRAME) then there will be another blank-screen gap before the next fixation cross appears indicating the start of the next pair.

(PLACE EXAMPLE PAIRS IN FRONT OF THE PARTICIPANT) Your task is to look out for the target pair, (POINT TO TARGET PAIR) which is this combination of patterns here. So whenever this pattern (POINT TO 'TARGET FIRST' IN TARGET PAIR) is followed by this pattern (POINT TO 'TARGET SECOND' IN TARGET PAIR) it indicates a target pair. We will call this pattern 'target first' (POINT TO 'TARGET FIRST' IN TARGET PAIR) and this pattern 'target second' (POINT TO TARGET SECOND). All other combinations of patterns are non-target pairs. Notice that non-target pairs could have something in common with the target pair: they could have a 'target first' (POINT TO TARGET FIRST IN TARGET PAIR AND THEN TARGET FIRST IN 1ST NON-TARGET PAIR) but a 'non-target second' (POINT TO NON-TARGET SECOND IN 1ST NON-TARGET PAIR); they could have a 'non-target first' (POINT TO NON-TARGET FIRST IN 2ND NON-TARGET PAIR) but a 'target second' (POINT TO TARGET SECOND IN TARGET PAIR AND THEN TARGET SECOND IN 2ND NON-TARGET PAIR); or they could have two non-targets (POINT TO NON-TARGET FIRST THEN NON-TARGET SECOND IN 3RD NON-TARGET PAIR).

DPX task rules. In this task, there are four rules you must follow. There are two buttons to press, either red (POINT TO RED KEY [N]) or green (POINT TO GREEN KEY [B]). The first rule is: respond to the first pattern of the pair by pressing the red

key if it is a target. The second rule is: respond to the first pattern by pressing the red key if it is a non-target. The third rule is: respond to the second pattern of the pair by pressing the green key if both patterns were targets. The fourth rule is: respond to the second pattern by pressing the red key if either or both patterns were non-targets.

DPX example. We will go through an example in detail now (ENSURE EXAMPLE RUN IS IN FRONT OF PARTICIPANT). When I point to a frame, I'd like you to tell me how you would respond. First you see a fixation cross indicating the start of the pair. (POINT TO 1ST PATTERN) Then you see the first pattern of the pair – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 1). Then there is a blank-screen delay. (POINT 2ND PATTERN) Then you will see the second pattern of the pair – what would you do here? (If 'press green') Good. (If 'press red') No you would press green (REPEAT RULE 3). Then there is a blank-screen gap before the next pair. In this next pair there is a fixation cross, (POINT TO 3RD PATTERN) then the first pattern – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 1). Then there is a delay, (POINT TO 4TH PATTERN) then the second pattern – what would you do here? (If 'press red') Good. (If 'press green') No you would press red (REPEAT RULE 4). (CONTINUE POINTING TO PATTERNS AND CONFIRMING RESPONSES, OR CORRCTING RESPONSES [AND REPEATING APPROPRIATE RULE IF NECESSARY], FOR REMAINDER OF EXAMPLE.)

Reconceptualisation of DPX task rules. I would like you to now think about the way in which the rules are represented in your *own mind*, which may be the same as, or different to, my description of the rules.

So I described the task as having four rules. (STATE ALL FOUR RULES.)

Here is a short list of possible instruction rules for the task (PLACE RULE SHEET IN FRONT OF PARTICIPANT). I'd like you to read this list and select the option that best reflects how the rules represented in your *own mind*. If none of the options are appropriate you can write your own description and suggest how many distinct rules there are at the bottom. Please take as long as you need to think carefully about the way in which the instructions are represented *in your own mind*.

Learning of task rules

I'll give you a chance to re-read the rule formats that you have chosen for each task so that you are clear on the rules before we start. (REMOVE RULE SHEET FROM PARTICIPANT'S VIEW) Now, without looking, would you repeat the rules of the task please? Remember that you have chosen N rules for the coloured shapes task and N rules for the dot pattern task. (If recalled without omissions) That's right (STATE ALL RULES IN THE SELECTED RULE FORMATS). (If recalled with omissions or in a different format to that chosen) That's not quite right (STATE THE APPROPRIATE RULE UNTIL RECELLED WITHOUT OMISSIONS, THEN, STATE ALL RULES IN THE SELECTED RULE FORMATS).

Task execution

Task execution (all 3 blocks). You will complete the task in three blocks. In the first block you will complete (if the first block of trials contains colour shape match trials) a long block of colour shape match trials followed by a long block of dot pattern trials/ (if the first block of trials contains DPX trials) a long block of dot pattern trials followed by a long block of colour shape match trials. In the next two blocks you will switch between the two tasks, but you will not know when.

Now remember, some of this will be very fast – try not to be too surprised and just do your best by responding as quickly and as accurately as you can. Press the space bar to start.

Participant states rules. Would you repeat the rules one last time please? (Whether recalled with or without omissions) Okay, thank you.